

日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

1c932 U.S. PTO
09/676644
09/29/00

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application:

1999年 9月30日

出願番号
Application Number:

平成11年特許願第277745号

出願人
Applicant(s):

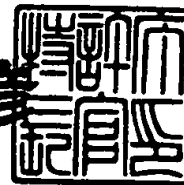
ソニー株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年 6月29日

特許庁長官
Commissioner,
Patent Office

近藤隆彦



出証番号 出証特2000-3052266

【書類名】 特許願

【整理番号】 9900749803

【提出日】 平成11年 9月30日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/20
G10L 3/00

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 本田 等

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 表 雅則

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 小川 浩明

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 包 洪長

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代表者】 出井 伸之

【代理人】

【識別番号】 100082131

【弁理士】

【氏名又は名称】 稲本 義雄

【電話番号】 03-3369-6479

【手数料の表示】

【予納台帳番号】 032089

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9708842

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声処理装置および音声処理方法、並びに記録媒体

【特許請求の範囲】

【請求項 1】 音声を音声認識する音声認識手段と、

前記音声認識手段による音声認識結果を対象として自然言語処理を行う自然言語処理手段と

を備える音声処理装置であって、

前記自然言語処理手段は、前記音声認識結果を対象とする自然言語処理によって得られる情報を、前記音声認識手段に対してフィードバックするフィードバック手段を含み、

前記音声認識手段は、前記フィードバック手段からフィードバックされる情報に基づいて処理を行う処理手段を含む

ことを特徴とする音声処理装置。

【請求項 2】 前記音声認識手段は、前記音声を、モデルを用いて認識し、

前記処理手段は、前記フィードバック手段からフィードバックされる情報に基づいて、前記モデルの適応を行う

ことを特徴とする請求項 1 に記載の音声処理装置。

【請求項 3】 前記フィードバック手段は、前記モデルの適応に利用すべき音声認識結果の区間、または利用すべきでない音声認識結果の区間のうちの少なくとも一方をフィードバックする

ことを特徴とする請求項 2 に記載の音声処理装置。

【請求項 4】 前記フィードバック手段は、確からしい前記音声認識結果をフィードバックする

ことを特徴とする請求項 2 に記載の音声処理装置。

【請求項 5】 前記フィードバック手段は、前記音声認識結果の信頼性をフィードバックする

ことを特徴とする請求項 2 に記載の音声処理装置。

【請求項 6】 前記フィードバック手段は、前記音声認識結果の内容が属するタスクをフィードバックする

ことを特徴とする請求項 2 に記載の音声処理装置。

【請求項 7】 前記フィードバック手段は、
前記モデルの適応に利用すべき音声認識結果の区間、若しくは利用すべきでない音声認識結果の区間のうちの少なくとも一方、
確からしい前記音声認識結果、
前記音声認識結果の信頼性、
または前記音声認識結果の内容が属するタスク
のうちの 1 以上をフィードバックする
ことを特徴とする請求項 2 に記載の音声処理装置。

【請求項 8】 前記音声を音声認識する音声認識ステップと、
前記音声認識ステップによる音声認識結果を対象として自然言語処理を行う自然言語ステップと
を備える音声処理方法であって、
前記自然言語処理ステップは、前記音声認識結果を対象とする自然言語処理によって得られる情報を、前記音声認識ステップに対してフィードバックするフィードバックステップを含み、
前記音声認識ステップは、前記フィードバックステップからフィードバックされる情報に基づいて処理を行う処理ステップを含む
ことを特徴とする音声処理方法。

【請求項 9】 音声を、コンピュータに処理させるプログラムが記録されている記録媒体であって、
前記音声を音声認識する音声認識ステップと、
前記音声認識ステップによる音声認識結果を対象として自然言語処理を行う自然言語ステップと
を備えるプログラムが記録されている記録媒体であって、
前記自然言語処理ステップは、前記音声認識結果を対象とする自然言語処理によって得られる情報を、前記音声認識ステップに対してフィードバックするフィードバックステップを含み、
前記音声認識ステップは、前記フィードバックステップからフィードバックさ

れる情報に基づいて処理を行う処理ステップを含む
ことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声処理装置および音声処理方法、並びに記録媒体に関し、特に、例えば、音声認識に用いるモデルの適応を、容易かつ高精度に行うこと等ができるようにする音声処理装置および音声処理方法、並びに記録媒体に関する。

【0002】

【従来の技術】

音声認識のアルゴリズムとしては、例えば、HMM(Hidden Markov Model)法などの、入力された音声（入力音声）を、モデルを用いて認識するものが知られている。即ち、HMM法では、あらかじめ学習によって、遷移確率（ある状態から、他の状態に遷移する確率）や、出力確率（状態が遷移するとき、あるシンボルが出力される確率）で規定されるモデル（HMM）が求められ、そのモデルを用いて、入力音声が認識される。

【0003】

ところで、音声認識を行う場合において、その認識精度を向上させるために、入力音声を用いて、逐次的に、モデルの適応を行うオンライン適応処理が知られている。オンライン適応処理によれば、話者の発話量に応じて段階的に音響モデルの高精度化や、言語モデルのタスク適応化等を図ることができ、従って、オンライン適応処理は、認識率を向上させるために、非常に有効な手段の1つである。

【0004】

【発明が解決しようとする課題】

モデルの適応化の方法は、いわゆる教師として、真の正解を与えて行われる「教師あり」の方法と、教師として、正解であると推測されるデータ（従って、正解かどうかは分からないデータ）を与えて行われる「教師なし」の方法の2種類に大別される。

【0005】

従来、「教師なし」の方法としては、例えば、特開平11-85184号公報に記載されているもの等があり、音声認識装置において、その音声認識結果を教師として、入力音声を用いて、モデルの適応が行われる。従って、従来の「教師なし」の方法では、モデルの適応において、音声認識結果が正しいかどうかを、ユーザに確認しないため、ユーザの負担が少ない反面、教師として使用されるデータの信頼性が十分でないことがあり、モデルが話者に対して十分に適応しないことがある。

【0006】

一方、「教師あり」の方法としては、例えば、Q. Huo et al., "A study of on-line Quasi-Bayes adaptation for DCHMM-bases speech recognition", Proceedings of International Conference on Acoustics, Speech and Signal Processing 1996, pp.705-708に記載されているもの等があり、音声認識装置において、ユーザに対して、一定の発話を要求し、その要求によって得られるユーザの発話を用いて、モデルの適応が行われる。あるいは、また、音声認識装置において、音声認識結果が正しい（正解である）かどうかの確認を、ユーザにしてもらい、正しいことが確認された音声認識結果を用いて、モデルの適応が行われる。

【0007】

しかしながら、ユーザに、一定の発話を要求してモデルの適応を行う方法は、オンライン適応に不向きである。また、ユーザに、音声認識結果の確認をしてもらってモデルの適応を行う方法では、ユーザの負担が大になる。

【0008】

さらに、モデルの適応を行う方法としては、例えば、特開平10-198395号公報において、言語モデル、あるいは、言語モデルを作成するためのデータを、特定の分野や話題等のタスク別に用意し、タスク別の言語モデル等を組み合わせて、高精度のタスク適応言語モデルを、オフラインで作成する方法が記載されているが、この方法によって、オンライン適応を行う場合には、発話内容のタスクを推定する必要があり、従って、音声認識装置を単独で用いて行うのは困難である。

【0009】

本発明は、このような状況に鑑みてなされたものであり、例えば、音声認識に用いるモデルの適応を、ユーザに負担をかけることなく、かつ精度良く行うこと等ができるようにするものである。

【0010】

【課題を解決するための手段】

本発明の音声処理装置は、自然言語処理手段が、音声認識結果を対象とする自然言語処理によって得られる情報を、音声認識手段に対してフィードバックするフィードバック手段を含み、音声認識手段が、フィードバック手段からフィードバックされる情報に基づいて処理を行う処理手段を含むことを特徴とする。

【0011】

音声認識手段には、音声を用いて認識させ、処理手段には、フィードバック手段からフィードバックされる情報に基づいて、モデルの適応を行わせることができる。

【0012】

フィードバック手段には、モデルの適応に利用すべき音声認識結果の区間、または利用すべきでない音声認識結果の区間のうちの少なくとも一方をフィードバックさせることができる。また、フィードバック手段には、確からしい音声認識結果をフィードバックさせることができる。さらに、フィードバック手段には、音声認識結果の信頼性をフィードバックさせることができる。また、フィードバック手段には、音声認識結果の内容が属するタスクをフィードバックさせることができる。

【0013】

フィードバック手段には、モデルの適応に利用すべき音声認識結果の区間、若しくは利用すべきでない音声認識結果の区間のうちの少なくとも一方、確からしい音声認識結果、音声認識結果の信頼性、または音声認識結果の内容が属するタスクのうちの1以上をフィードバックさせることができる。

【0014】

本発明の音声処理方法は、自然言語処理ステップが、音声認識結果を対象とす

る自然言語処理によって得られる情報を、音声認識ステップに対してフィードバックするフィードバックステップを含み、音声認識ステップが、フィードバックステップからフィードバックされる情報に基づいて処理を行う処理ステップを含むことを特徴とする。

【0015】

本発明の記録媒体は、自然言語処理ステップが、音声認識結果を対象とする自然言語処理によって得られる情報を、音声認識ステップに対してフィードバックするフィードバックステップを含み、音声認識ステップが、フィードバックステップからフィードバックされる情報に基づいて処理を行う処理ステップを含むことを特徴とする。

【0016】

本発明の音声処理装置および音声処理方法、並びに記録媒体においては、音声認識結果を対象とする自然言語処理によって得られる情報がフィードバックされ、そのフィードバックされる情報に基づいて処理が行われる。

【0017】

【発明の実施の形態】

図1は、本発明を適用した音声処理システム（システムとは、複数の装置が論理的に集合した物をいい、各構成の装置が同一筐体中にあるか否かは問わない）の一実施の形態の構成例を示している。

【0018】

この音声処理システムでは、音声が入力されると、その音声に対する返答が出力されたり、また、その音声の対訳が出力されるようになっている。さらに、入力された音声に対する返答を、その音声の言語以外の言語に翻訳して出力することもできるようになっている。

【0019】

即ち、音声認識部1には、例えば日本語などによる音声が入力されるようになっており、音声認識部1は、入力された音声を音声認識し、その音声認識結果としてのテキスト、その他付随する情報を、機械翻訳部2や、表示部4、対話管理部5などに、必要に応じて出力する。

【 0 0 2 0 】

機械翻訳部 2 は、音声認識部 1 が出力する音声認識結果を解析し、入力された音声の言語以外の、例えば英語に機械翻訳し、その翻訳結果としてのテキスト、その他付随する情報を、音声合成部 3 や、表示部 4、対話管理部 5 などに、必要に応じて出力する。音声合成部 3 は、機械翻訳部 2 や対話管理部 5 などの出力に基づいて音声合成処理を行い、これにより、入力された音声に対する返答、あるいはその音声の、他の言語への翻訳結果としての合成音を出力する。

【 0 0 2 1 】

表示部 4 は、例えば、液晶ディスプレイ等で構成され、音声認識部 1 による音声認識結果や、機械翻訳部 2 による機械翻訳結果、対話管理部 5 が生成する返答等を、必要に応じて表示する。

【 0 0 2 2 】

対話管理部 5 は、音声認識部 1 の音声認識結果に対する応答を生成し、機械翻訳部 2 や、音声合成部 3、表示部 4、対話管理部 5 に、必要に応じて出力する。さらに、対話管理部 5 は、機械翻訳部 2 の機械翻訳結果に対する応答を生成し、音声合成部 3 や、表示部 4 に、必要に応じて出力する。

【 0 0 2 3 】

以上のように構成される音声処理システムにおいて、入力された音声に対する返答を出力する場合には、まず、その入力された音声は、音声認識部 1 で音声認識され、対話管理部 5 に出力される。対話管理部 5 では、音声認識部 1 による音声認識結果に対する応答が生成され、音声合成部 3 に供給される。音声合成部 3 では、対話管理部 5 からの応答に対応する合成音が生成されて出力される。

【 0 0 2 4 】

また、入力された音声の対訳を出力する場合には、その入力された音声は、音声認識部 1 で音声認識され、機械翻訳部 2 に供給される。機械翻訳部 2 では、音声認識部 1 による音声認識結果が機械翻訳され、音声合成部 3 に供給される。音声合成部 3 では、機械翻訳部 2 からの翻訳結果に対応する合成音が生成されて出力される。

【 0 0 2 5 】

さらに、入力された音声に対する返答を、その音声の言語以外の言語に翻訳して出力する場合には、まず、その入力された音声は、音声認識部 1 で音声認識され、対話管理部 5 に出力される。対話管理部 5 では、音声認識部 1 による音声認識結果に対する返答が生成され、機械翻訳部 2 に供給される。機械翻訳部 2 では、対話管理部 5 からの返答が機械翻訳され、音声合成部 3 に供給される。音声合成部 3 では、機械翻訳部 2 からの翻訳結果に対応する合成音が生成されて出力される。

【 0 0 2 6 】

なお、入力された音声に対する返答を、その音声の言語以外の言語に翻訳して出力する場合には、音声認識部 1 による音声認識結果を、機械翻訳部 2 で機械翻訳し、その翻訳結果に対する応答を、対話管理部 5 で生成し、合成音で出力するようにすることも可能である。

【 0 0 2 7 】

次に、図 1 の音声処理システムでは、図 2 に示すように、ユーザの発話（入力音声）が、音声認識部 1 において音声認識され、その音声認識結果が、機械翻訳や対話管理等の自然言語処理を行う自然言語処理部としての機械翻訳部 2 や対話管理部 5 で処理されるが、機械翻訳部 2 や対話管理部 5 では、音声認識結果を対象とする自然言語処理によって得られる情報が、音声認識部 1 に対してフィードバックされ、音声認識部 1 では、このようにしてフィードバックされる情報（以下、適宜、フィードバック情報という）に基づいて各種の処理が行われるようになっている。

【 0 0 2 8 】

具体的には、例えば、機械翻訳部 2 や対話管理部 5 では、音声認識部 1 で用いられるモデルの適応を行うのに有用な情報がフィードバックされ、音声認識部 1 では、その情報に基づいて、モデルの適応が行われる。また、例えば、機械翻訳部 2 や対話管理部 5 では、音声認識部 1 による音声認識結果の自然言語処理を行い易いように、その構成単位を変更するための情報がフィードバックされ、音声認識部 1 では、その情報に基づいて、音声認識結果の構成単位を変更するための処理が行われる。さらに、例えば、機械翻訳部 2 や対話管理部 5 では、音声認識

部 1 による音声認識結果の誤りを修正等するための情報がフィードバックされ、音声認識部 1 では、その情報に基づいて、正しい音声認識結果を得るための処理が行われる。

【0029】

次に、図 3 は、図 1 の音声認識部 1 の第 1 の構成例を示している。

【0030】

ユーザの発話は、マイク 11 に入力され、マイク 11 では、その発話が、電気信号としての音声信号に変換される。この音声信号は、A/D (Analog Digital) 変換部 12 に供給される。A/D 変換部 12 では、マイク 11 からのアナログ信号である音声信号がサンプリング、量子化され、ディジタル信号である音声データに変換される。この音声データは、特徴抽出部 13 に供給される。

【0031】

特徴抽出部 13 は、A/D 変換部 12 からの音声データについて、適当なフレームごとに、例えば、スペクトルや、線形予測係数、ケプストラム係数、線スペクトル対等の特徴パラメータを抽出し、特徴量バッファ 14 およびマッチング部 15 に供給する。特徴量バッファ 14 では、特徴抽出部 13 からの特徴パラメータが一時記憶される。

【0032】

マッチング部 15 は、特徴抽出部 13 からの特徴パラメータ、または特徴量バッファ 14 に記憶された特徴パラメータに基づき、音響モデルデータベース 16、辞書データベース 17、および文法データベース 18 を必要に応じて参照しながら、マイク 11 に入力された音声（入力音声）を認識する。

【0033】

即ち、音響モデルデータベース 16 は、音声認識する音声の言語における個々の音素や音節などの音響的な特徴を表す音響モデルを記憶している。ここで、音響モデルとしては、例えば、HMM (Hidden Markov Model) などを用いることができる。辞書データベース 17 は、認識対象の各単語について、その発音に関する言語モデルが記述された単語辞書を記憶している。文法データベース 18 は、辞書データベース 17 の単語辞書に登録されている各単語が、どのように連鎖す

る（つながる）かを記述した文法規則を記憶している。ここで、文法規則としては、例えば、文脈自由文法（CFG）や、統計的な単語連鎖確率（N-gram）などに基づく規則を用いることができる。

【0034】

マッチング部15は、辞書データベース17の単語辞書を参照することにより、音響モデルデータベース16に記憶されている音響モデルを接続することで、単語の音響モデル（単語モデル）を構成する。さらに、マッチング部15は、幾つかの単語モデルを、文法データベース18に記憶された文法規則を参照することにより接続し、そのようにして接続された単語モデルを用いて、特徴パラメータに基づき、例えば、HMM法等によって、マイク11に入力された音声を認識する。

【0035】

そして、マッチング部15による音声認識結果は、例えば、テキスト等で出力される。

【0036】

一方、適応処理部19は、マッチング部15による音声認識結果を受信しており、例えば、対話管理部5等から、後述するようなフィードバック情報を受信すると、そのフィードバック情報に基づいて、マッチング部15による音声認識結果から、音響モデルデータベース16の音響モデルや、辞書データベース17の言語モデルの適応を行うのに適切なものを得て（精度の良い適応を行うのに、教師として用いるべき音声認識結果を得て）、その音声認識結果を教師として、音響モデルデータベース16の音響モデルや、辞書データベース17の言語モデル（以下、適宜、両方含めて、単に、モデルという）のオンライン適応を行う。

【0037】

ここで、例えば、音響モデルとして、HMMが採用されている場合には、適応処理部19では、そのHMMを表す遷移確率、あるいは出力確率を規定する平均値および分散等のモデルのパラメータを、音声認識結果を用いて変更することで、モデルの適応が行われる。

【0038】

次に、図 4 は、図 1 の機械翻訳部 2 の第 1 の構成例を示している。

【 0 0 3 9 】

テキスト解析部 2 1 には、音声認識部 1 が出力する音声認識結果としてのテキストや、対話管理部 5 が出力する返答としてのテキストが、機械翻訳の対象として入力されるようになっており、テキスト解析部 2 1 は、辞書データベース 2 4 や解析用文法データベース 2 5 を参照しながら、そのテキストを解析する。

【 0 0 4 0 】

即ち、辞書データベース 2 4 には、各単語の表記や、解析用文法を適用するために必要な品詞情報などが記述された単語辞書が記憶されている。また、解析用文法データベース 2 5 には、単語辞書に記述された各単語の情報に基づいて、単語連鎖に関する制約等が記述された解析用文法規則が記憶されている。そして、テキスト解析部 2 1 は、その単語辞書や解析用文法規則に基づいて、そこに入力されるテキスト（入力テキスト）の形態素解析や、構文解析等を行い、その入力テキストを構成する単語や構文の情報等の言語情報を抽出する。ここで、テキスト解析部 2 1 における解析方法としては、例えば、正規文法や、文脈自由文法、統計的な単語連鎖確率を用いたものなどがある。

【 0 0 4 1 】

テキスト解析部 2 1 で得られた入力テキストの解析結果としての言語情報は、言語変換部 2 2 に供給される。言語変換部 2 2 は、言語変換データベース 2 6 を参照し、入力テキストの言語の言語情報を、翻訳結果の言語の言語情報に変換する。

【 0 0 4 2 】

即ち、言語変換データベース 2 6 には、入力言語（言語変換部 2 2 への入力の言語）の言語情報から、出力言語（言語変換部 2 2 からの出力の言語）の言語情報への変換パターン（テンプレート）や、入力言語と出力言語との対訳用例およびその対訳用例と入力言語との間の類似度の計算に用いられるシソーラス等の、言語情報を変換するための言語変換データが記憶されている。そして、言語変換部 2 2 では、このような言語変換データに基づいて、入力テキストの言語の言語情報が、出力言語の言語情報に変換される。

【 0 0 4 3 】

言語変換部 2 2 で得られた出力言語の言語情報は、テキスト生成部 2 3 に供給され、テキスト生成部 2 3 は、辞書データベース 2 7 および生成用文法データベース 2 8 を参照することにより、出力言語の言語情報から、入力テキストを出力言語に翻訳したテキストを生成する。

【 0 0 4 4 】

即ち、辞書データベース 2 7 には、出力言語の文を生成するのに必要な単語の品詞や活用形等の情報が記述された単語辞書が記憶されており、また、生成用文法データベース 2 8 には、出力言語の文を生成するのに必要な単語の活用規則や語順の制約等の生成用文法規則が記憶されている。そして、テキスト生成部 2 3 は、これらの単語辞書および生成用文法規則に基づいて、言語変換部 2 2 からの言語情報を、テキストに変換して出力する。

【 0 0 4 5 】

次に、図 5 は、図 1 の音声合成部 3 の構成例を示している。

【 0 0 4 6 】

テキスト解析部 3 1 には、機械翻訳部 2 が出力する翻訳結果としてのテキストや、対話管理部 5 が出力する返答としてのテキストが、音声合成処理の対象として入力されるようになっており、テキスト解析部 3 1 は、辞書データベース 3 4 や解析用文法データベース 3 5 を参照しながら、そのテキストを解析する。

【 0 0 4 7 】

即ち、辞書データベース 3 4 には、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、解析用文法データベース 3 5 には、辞書データベース 3 4 の単語辞書に記述された単語について、単語連鎖に関する制約等の解析用文法規則が記憶されている。そして、テキスト解析部 3 1 は、この単語辞書および解析用文法規則に基づいて、そこに入力されるテキストの形態素解析や構文解析等の解析を行い、後段の規則合成部 3 2 で行われる規則音声合成に必要な情報を抽出する。ここで、規則音声合成に必要な情報としては、例えば、ポーズの位置や、アクセントおよびイントネーションを制御するための情報その他の韻律情報や、各単語の発音等の音韻情報などがある。

【 0 0 4 8 】

テキスト解析部 3 1 で得られた情報は、規則合成部 3 2 に供給され、規則合成部 3 2 では、音素片データベース 3 6 を用いて、テキスト解析部 3 1 に入力されたテキストに対応する合成音の音声データ（デジタルデータ）が生成される。

【 0 0 4 9 】

即ち、音素片データベース 3 6 には、例えば、C V (Consonant, Vowel) や、V C V、C V C 等の形で音素片データが記憶されており、規則合成部 3 2 は、テキスト解析部 3 1 からの情報に基づいて、必要な音素片データを接続し、さらに、ポーズ、アクセント、イントネーション等を適切に付加することで、テキスト解析部 3 1 に入力されたテキストに対応する合成音の音声データを生成する。

【 0 0 5 0 】

この音声データは、D A 変換部 3 3 に供給され、そこで、アナログ信号としての音声信号に変換される。この音声信号は、図示せぬスピーカに供給され、これにより、テキスト解析部 3 1 に入力されたテキストに対応する合成音が出力される。

【 0 0 5 1 】

次に、図 6 は、図 1 の対話管理部 5 の第 1 の構成例を示している。

【 0 0 5 2 】

例えば、音声認識部 1 による音声認識結果は、対話処理部 4 1 および音声区間検出部 4 2 に供給されるようになっており、対話処理部 4 1 は、その音声認識結果に対する返答を生成して出力する。一方、音声区間検出部 4 2 は、対話処理部 4 1 が出力する返答を監視しており、その返答に基づいて、音声認識部 1 が出力する音声認識結果の区間から、モデルの適応に利用すべき部分の区間（以下、適宜、適応用区間という）を検出し、その適応用区間を、フィードバック情報として、音声認識部 1（の適応処理部 1 9）にフィードバックする。

【 0 0 5 3 】

なお、ここでは、音声区間検出部 4 2 において、モデルの適応に利用すべき音声認識結果の区間を、フィードバック情報として、音声認識部 1 にフィードバックするようにしたが、その他、例えば、モデルの適応に利用すべきでない音声認

識結果の区間や、モデルの適応に利用すべき音声認識結果の区間と、利用すべきでない音声認識結果の区間の両方をフィードバックするようにしても良い。

【0054】

次に、図7のフローチャートを参照して、音声認識部1が図3に示したように、対話管理部5が図6に示したように、それぞれ構成される場合の、図1の音声処理システムの動作について説明する。

【0055】

ユーザが発話を行い、音声認識部1に対して、そのユーザの発話である音声が入力されると、音声認識部1では、ステップS1において、その音声が発話認識され、その音声認識結果としてのテキストが、例えば、対話管理部5に出力される。

【0056】

対話管理部5では、ステップ2において、対話処理部41が、音声認識部1からの音声認識結果に対する返答を生成する対話処理を行い、その返答を出力する。そして、ステップS3に進み、音声区間検出部42は、対話処理部41が出力した返答から、音声認識部1からの音声認識結果が正しいかどうかを判定する。ステップS3において、音声認識部1からの音声認識結果が正しくないと判定された場合、ステップS4およびS5をスキップして、処理を終了する。

【0057】

一方、ステップS3において、音声認識部1からの音声認識結果が正しいと判定された場合、音声区間検出部42は、音声認識部1からの音声認識結果から、正しい部分を検出し、その部分の区間を、適応用区間として、音声認識部1（図3）の適応処理部19に送信する。

【0058】

適応処理部19では、ステップS5において、マッチング部15が出力する音声認識結果のうち、音声区間検出部42が出力する適応用区間の部分のみを用いて、モデルの適応が行われ、処理を終了する。

【0059】

以上の処理によれば、音声認識に用いるモデルの適応を、ユーザに負担をかけ

ることなく、かつ精度良く行うことができる。

【0060】

即ち、例えば、いま、図1の音声処理システムと、ユーザとの間で、以下のよう
な、コンサートのチケットの購入に関する対話が行われたとする。

【0061】

ユーザ：「こんにちは。9月11日のベルリンフィルのチケットがほしいんだけ
ど。」

．．．（1）

返答：「9月11日、ベルリンフィルのチケットですね。S席からD席まであり
ますが、何席に致しましょうか？」

．．．（2）

ユーザ：「S席で。」

．．．（3）

返答：「A席ですね。」

．．．（4）

ユーザ：「いや、S席。」

．．．（5）

返答：「S席ですね。1階の4列目24番の席をお取り致します。2万8千円で
すが、よろしいでしょうか？」

．．．（6）

ユーザ：「それでお願い。」

．．．（7）

返答：「かしこまりました。」

．．．（8）

【0062】

（1）乃至（8）の対話において、音声区間検出部42では、ユーザの発話（
1）、（5）、（7）の音声認識結果は、それぞれに対する返答（2）、（6）
、（8）から正しいと判定される。一方、ユーザの発話（3）の音声認識結果は
、それに対する返答（4）に対して、ユーザが言い直しの発話（5）を行ってい

るので、誤っていると判定される。

【0063】

この場合、音声区間検出部42は、ユーザの発話(1)、(5)、(7)についての、正しい音声認識結果の区間を、フィードバック情報として、適応処理部19にフィードバックし(誤っていると判定されたユーザの発話(3)の音声認識結果の区間はフィードバックされない)、その結果、適応処理部19では、この正しい音声認識結果の区間だけを用いて、即ち、その正しい音声認識結果を教師とするとともに、その音声認識結果に対応する音声区間のユーザの発話を生徒として、モデルの適応が行われる。

【0064】

従って、誤った音声認識結果を用いずに、正しい音声認識結果だけを用いて、モデルの適応が行われるので、精度の良い(認識率を向上させるように)、モデルの適応を行うことができる。さらに、この場合、ユーザに負担をかけることもない。

【0065】

次に、図8は、図1の対話管理部5の第2の構成例を示している。なお、図中、図6における場合と対応する部分については、同一の符号を付してある。即ち、図8の対話管理部5は、音声区間検出部42に替えて、修正部43が設けられている他は、図6における場合と同様に構成されている。

【0066】

修正部43は、対話処理部41が出力する返答を監視しており、その返答に基づいて、音声認識部1が出力する音声認識結果が正しいかどうかを判定し、正しい場合には、その音声認識結果を、そのままフィードバック情報として、適応処理部19にフィードバックする。また、修正部43は、音声認識部1が出力する音声認識結果が誤っている場合には、その音声認識結果を正しく(確からしく)修正し、フィードバック情報として、適応処理部19にフィードバックする。

【0067】

次に、図9のフローチャートを参照して、音声認識部1が図3に示したように、対話管理部5が図8に示したように、それぞれ構成される場合の、図1の音声

処理システムの動作について説明する。

【0068】

ステップS11またはS12では、図7のステップS1またはS2における場合とそれぞれ同様の処理が行われ、これにより、対話処理部41から、音声認識部1による音声認識結果に対する返答が出力される。

【0069】

そして、ステップS13に進み、修正部43は、対話処理部41が出力した返答から、音声認識部1からの音声認識結果が正しいかどうかを判定する。ステップS13において、音声認識部1からの音声認識結果が正しいと判定された場合、ステップS14に進み、修正部43は、その正しい音声認識結果を、フィードバック情報として、音声認識部1の適応処理部19に送信する。

【0070】

適応処理部19では、ステップS15において、修正部43からのフィードバック情報としての正しい音声認識結果を用いて、モデルの適応が行われ、処理を終了する。

【0071】

一方、ステップS13において、音声認識部1からの音声認識結果が誤っていると判定された場合、修正部43は、音声認識部1からの音声認識結果を、対話処理部41からの返答に基づいて正しく（確からしく）修正し、その修正した音声認識結果を、フィードバック情報として、音声認識部1の適応処理部19に送信する。

【0072】

適応処理部19では、ステップS15において、修正部43からの修正後の音声認識結果、即ち、正しい（確からしい）音声認識結果を用いて、モデルの適応が行われ、処理を終了する。

【0073】

以上の処理によっても、音声認識に用いるモデルの適応を、ユーザに負担をかけることなく、かつ精度良く行うことができる。

【0074】

即ち、例えば、いま、音声処理システムとユーザとの間で、上述の（１）乃至（８）の対話が行われたとすると、修正部４３では、ユーザの発話（１）、（５）、（７）の音声認識結果は、それぞれに対する返答（２）、（６）、（８）から正しいと判定される。一方、ユーザの発話（３）の音声認識結果は、それに対する返答（４）に対して、ユーザが言い直しの発話（５）を行っているので、誤っていると判定される。

【００７５】

この場合、修正部４３は、ユーザの発話（１）、（５）、（７）についての正しい音声認識結果を、そのまま、フィードバック情報として、適応処理部１９にフィードバックし、適応処理部１９では、この正しい音声認識結果と、その音声認識結果に対するユーザの発話（１）、（５）、（７）とを用いて、モデルの適応が行われる。

【００７６】

さらに、修正部４３は、ユーザの発話（３）についての誤った音声認識結果を、その後のユーザの発話（５）についての正しい音声認識結果に基づいて修正する。即ち、修正部４３では、ユーザの発話（３）「Ｓ席で。」に対する返答（４）「Ａ席ですね。」について、その後のユーザの発話（５）「いや、Ｓ席で。」の正しい音声認識結果から、ユーザの発話（３）における「Ｓ」を、「Ａ」に誤認識したことと、その音声認識結果として、「Ｓ」が正しいことが解析される。そして、修正部４３は、その解析結果に基づいて、ユーザの発話（３）における「Ｓ」を「Ａ」に誤認識した音声認識結果を正しく修正し、その修正後の音声認識結果を、フィードバック情報として、適応処理部１９にフィードバックする。この場合、適応処理部１９では、その修正後の音声認識結果と、その音声認識結果に対するユーザの発話（３）とを用いて、モデルの適応が行われる。

【００７７】

従って、音声認識結果が誤っていても、それを、正しい（確からしい）音声認識結果に修正して、モデルの適応が行われるので、精度の良い（認識率を向上させるように）、モデルの適応を行うことができる。さらに、この場合、ユーザに負担をかけることもない。

【 0 0 7 8 】

次に、図 1 0 は、図 1 の対話管理部 5 の第 3 の構成例を示している。なお、図中、図 6 における場合と対応する部分については、同一の符号を付してある。即ち、図 1 0 の対話管理部 5 は、音声区間検出部 4 2 に替えて、信頼度計算部 4 4 が設けられている他は、図 6 における場合と同様に構成されている。

【 0 0 7 9 】

信頼度計算部 4 4 は、対話処理部 4 1 が出力する返答を監視しており、その返答に基づいて、音声認識部 1 が出力する音声認識結果の信頼性を表す信頼度を算出し、その信頼度を、フォードバック情報として、適応処理部 1 9 にフィードバックする。

【 0 0 8 0 】

次に、図 1 1 のフローチャートを参照して、音声認識部 1 が図 3 に示したように、対話管理部 5 が図 1 0 に示したように、それぞれ構成される場合の、図 1 の音声処理システムの動作について説明する。

【 0 0 8 1 】

ステップ S 2 1 または S 2 2 では、図 7 のステップ S 1 または S 2 における場合とそれぞれ同様の処理が行われ、これにより、対話処理部 4 1 から、音声認識部 1 による音声認識結果に対する返答が出力される。

【 0 0 8 2 】

そして、信頼度計算部 4 4 は、ステップ S 2 3 において、対話処理部 4 1 が出力した返答から、音声認識部 1 からの音声認識結果の信頼度としての、例えば、0 乃至 1 の範囲の実数値を計算し、ステップ S 2 4 に進み、その信頼度を、フィードバック情報として、音声認識部 1 の適応処理部 1 9 に送信する。

【 0 0 8 3 】

適応処理部 1 9 では、ステップ S 2 5 において、信頼度計算部 4 4 からのフィードバック情報としての信頼度を用いて、モデルの適応が行われ、処理を終了する。

【 0 0 8 4 】

以上の処理によっても、音声認識に用いるモデルの適応を、ユーザに負担をか

けることなく、かつ精度良く行うことができる。

【0085】

即ち、例えば、いま、音声処理システムとユーザとの間で、上述の(1)乃至(8)の対話が行われたとすると、信頼度計算部44では、ユーザの発話(1)、(5)、(7)の音声認識結果は、それぞれに対する返答(2)、(6)、(8)から正しいと判定される。一方、ユーザの発話(3)の音声認識結果は、それに対する返答(4)に対して、ユーザが言い直しの発話(5)を行っているので、誤っていると判定される。

【0086】

この場合、信頼度計算部44は、例えば、ユーザの発話(1)、(5)、(7)についての正しい音声認識結果の信頼度を、1として、また、ユーザの発話(3)についての誤った音声認識結果の信頼度を、0として、それぞれ、適応処理部19にフィードバックする。そして、適応処理部19では、ユーザの発話(1)、(3)、(5)、(7)と、その音声認識結果とを、それぞれの音声認識結果の信頼度に応じた重みで用いて、モデルの適応が行われる。

【0087】

従って、この場合、実質的に、正しい音声認識結果だけを用いて、モデルの適応が行われるので、精度の良い、モデルの適応を行うことができる。さらに、この場合、ユーザに負担をかけることもない。

【0088】

なお、信頼度としては、0と1の間の中間値を用いることも可能であり、そのような中間値の信頼度の算出は、例えば、音声認識部1による音声認識結果の尤度を用いて行うことが可能である。さらに、そのような信頼度を用いたモデルの適応は、例えば、次式にしたがって行うことが可能である。

【0089】

$$P_{\text{new}} = (1 - (1 - \alpha) \times R) \times P_{\text{old}} + (1 - \alpha) \times R \times P_{\text{adapt}}$$

但し、 P_{new} は、適応後のモデルのパラメータ（前述したように、モデルがHMMであれば、遷移確率や、出力確率を規定する平均値および分散など）を表し、 α は、適応を行うための所定の定数を表す。また、 R は、信頼度を表し、 P_{old}

は、適応前のモデルのパラメータを表す。さらに、 P_{adapt} は、適応に用いるユーザの発話から得られるデータを表す。

【0090】

次に、図12は、図1の対話管理部5の第4の構成例を示している。なお、図中、図6における場合と対応する部分については、同一の符号を付してある。即ち、図12の対話管理部5は、音声区間検出部42に替えて、タスク推定部45が設けられている他は、図6における場合と同様に構成されている。

【0091】

タスク推定部45は、対話処理部41が出力する返答を監視しており、その返答に基づいて、音声認識部1が出力する音声認識結果の内容が属するタスクを推定し、そのタスクを、フィードバック情報として、適応処理部19にフィードバックする。

【0092】

次に、図13のフローチャートを参照して、音声認識部1が図3に示したように、対話管理部5が図10に示したように、それぞれ構成される場合の、図1の音声処理システムの動作について説明する。なお、ここでは、音声認識部1の辞書データベース17において、例えば、コンサートチケットの予約用の言語モデルや、ホテルの予約用の言語モデル、航空券の予約用の言語モデル、新聞の読み上げ等のディクテーション用の言語モデル、その他各種の言語モデルが各タスク別に登録された辞書が記憶されているものとする。

【0093】

この場合、ステップS31またはS32において、図7のステップS1またはS2における場合とそれぞれ同様の処理が行われ、これにより、対話処理部41から、音声認識部1による音声認識結果に対する返答が出力される。

【0094】

そして、タスク推定部45は、ステップS33において、対話処理部41が出力した返答から、音声認識部1からの音声認識結果の内容が属するタスク（その内容が、どのような分野、あるいは話題であるのか）を認識し、ステップS34に進み、そのタスクを、フィードバック情報として、音声認識部1の適応処理部

19に送信する。

【0095】

適応処理部19では、ステップS35において、タスク推定部45からのフィードバック情報としてのタスクを用いて、モデルの適応が行われ、処理を終了する。

【0096】

即ち、例えば、いま、音声処理システムとユーザとの間で、上述の(1)乃至(8)の対話が行われたとすると、タスク推定部45では、ユーザの発話の音声認識結果と、それに対する返答とから、タスクが、コンサートチケットの予約に関するものであることが推定され、適応処理部19にフィードバックされる。この場合、適応処理部19では、辞書データベース17におけるタスク別の言語モデルのうち、コンサートチケットの予約用の言語モデルのみを対象とした適応が行われる。

【0097】

従って、音声認識に用いるモデルの適応を、ユーザに負担をかけることなく、かつ精度良く行うこと等ができる。

【0098】

なお、辞書データベース17には、言語モデルそのものではなく、言語モデルを作成するのに用いるデータを、タスク別に記憶させておくようにし、そのような辞書データベース17を対象として、適応を行うことも可能である。

【0099】

また、上述の場合には、タスク別に用意された言語モデルの適応を行うようにしたが、その他、例えば、音響モデルがタスク別に用意されている場合には、そのタスク別の音響モデルの適応を行うようにすることも可能である。

【0100】

即ち、音響モデルについては、数字の認識率を高めるために、数字用の音響モデル（以下、適宜、数字モデルという）が、数字以外の音響モデル（以下、適宜、一般の音響モデルという）とは別に用意されることがある。ここで、数字モデルと一般の音響モデルとを分けて行う音声認識については、例えば、KDD研究

所、河井 恒、電子情報通信学会研究報告SP98-69等に、その詳細が記載されている。

【0101】

このように、数字モデルと一般の音響モデルとが分けて用意されている場合には、タスク推定部45において、音声認識結果のタスクが数字がどうかを推定し、適応処理部19において、その推定結果を用いて、数字モデルや、一般の音響モデルの適応を行うようにすることができる。

【0102】

即ち、例えば、いま、音声処理システムとユーザとの間で、上述の(1)乃至(8)の対話が行われたとすると、タスク推定部45では、ユーザの発話の音声認識結果と、それに対する返答とから、ユーザの発話(1)「こんにちは。9月11日のベルリンフィルのチケットがほしいんだけど。」の音声認識結果における「9」と「1」は、数字のタスクとして推定され、適応処理部19にフィードバックされる。この場合、適応処理部19では、数字モデルの適応は、ユーザの発話およびその音声認識結果のうちの「9」と「1」の部分を用いて行われ、一般の音響モデルの適応は、ユーザの発話およびその音声認識結果のうちの他の部分を用いて行われる。

【0103】

なお、モデルの適応は、図6乃至図13を参照して説明した4つの適応の方法うちの2以上を組み合わせることも可能である。

【0104】

また、音声認識結果の翻訳を行う場合には、上述のようなフィードバック情報は、機械翻訳部2から、音声認識部1に対して出力させるようにすることが可能である。

【0105】

次に、図14は、図1の音声認識部1の第2の構成例を示している。なお、図中、図3における場合と対応する部分については、同一の符号を付してある。即ち、図14の音声認識部1は、適応処理部19が設けられていないことを除けば、図3における場合と基本的に同様に構成されている。

【0106】

但し、図14の実施の形態においては、マッチング部15は、例えば、機械翻訳部2から、後述する変更信号を、フィードバック情報として受信することができるようになっており、変更信号を受信した場合、音声認識結果を構成する構成単位を変更して出力するようになっている。

【0107】

即ち、マッチング部15は、例えば、入力音声「これは何ですか」に対して、音声認識結果「これは何ですか」が得られた場合において、変更信号を受信したとき、1つの構成単位「これは何ですか」でなる音声認識結果を、3つの構成単位「これは」、「何です」、「か」でなる音声認識結果や、5つの構成単位「これ」、「は」、「何」、「です」、「か」でなる音声認識結果に変更して出力し直すようになっている。

【0108】

このような構成単位の変更は、最初に得られた音声認識結果「これは何ですか」を構成する文節や単語を切り離すことで行うこともできるし、マッチング部15による音声認識の処理単位を変更することで行うこともできる。

【0109】

マッチング部15による音声認識の処理単位を変更することにより、その音声認識結果の構成単位を変更する場合においては、その変更のための情報を、辞書データベース17の単語辞書や、文法データベース18の文法規則に記述しておくことを行うことが可能である。

【0110】

即ち、例えば、単語辞書に、文節「これは」と、その文節を構成する単語（形態素）「これ」および「は」とを対応付けて記述しておく。この場合、マッチング部15では、単語辞書を参照することにより、入力音声「これは」に対して、「これは」という1つの構成単位で構成される音声認識結果を得ることも可能となるし、「これ」と「は」との2つの構成単位で構成される音声認識結果を得ることも可能となる。

【0111】

なお、上述の場合には、文節と、その文節を構成する単語とを対応付けておくようにしたが、文と、その文を構成する文節とを対応付けておくことも可能であるし、さらに、その文節を構成する単語も対応付けておくことも可能である。

【0112】

一方、文法データベース18の文法規則によって、マッチング部15による音声認識の処理単位を変更する場合には、例えば、主格が、代名詞と助詞「は」とを接続して構成されることを、文法規則に記述しておく。この場合も、マッチング部15では、文法規則を参照することにより、代名詞「これ」と助詞「は」とからなる主格を表す入力音声「これは」に対して、「これは」という1つの構成単位で構成される音声認識結果を得ることも可能となるし、「これ」と「は」との2つの構成単位で構成される音声認識結果を得ることも可能となる。

【0113】

ここで、マッチング部15による音声認識の処理単位を変更は、単語辞書または文法規則のいずれか一方に基づいて行う他、その両方に基づいて行うことも可能である。さらに、マッチング部15による音声認識の処理単位を変更は、単語辞書を複数用意するとともに、各単語辞書に対応する文法規則も用意し、変更信号に基づいて、音声認識に用いる単語辞書と文法規則の組合せを選択するようにすることで行うことも可能である。

【0114】

なお、マッチング部15による音声認識の処理単位を変更することにより、音声認識結果の構成単位を変更する場合には、マッチング部15では、特徴量バッファ14に記憶された特徴パラメータを用いて、再度処理が行われる。

【0115】

次に、図15は、音声認識部1が図14に示すように構成される場合の、図1の機械翻訳部2の構成例（第2の構成例）を示している。なお、図中、図4における場合と対応する部分については、同一の符号を付してある。即ち、図15の機械翻訳部2は、図4における場合と基本的に同様に構成されている。

【0116】

但し、図15の実施の形態においては、テキスト解析部21は、入力テキスト

の構成単位が、その入力テキストの解析が可能な単位（入力テキストの解析を行うのに適切な単位）であるかどうかを判定し、可能な単位であれば、上述したように、入力テキストの解析を行う。一方、入力テキストの構成単位が、その入力テキストの解析が可能な単位でない場合には、テキスト解析部 2 1 は、入力テキストの構成単位の変更を指示する変更信号を、フィードバック情報として、音声認識部 1 に送信する。音声認識部 1 では、上述したように、変更信号に基づいて、入力テキストとなる音声認識結果の構成単位が変更され、その結果、テキスト解析部 2 1 には、構成単位が変更された音声認識結果が、入力テキストとして供給される。そして、テキスト解析部 2 1 では、再び、その入力テキストの構成単位が解析可能な単位であるかどうかを判定され、以下、同様の処理が繰り返される。

【0 1 1 7】

ここで、対話管理部 5 も、機械翻訳部 2 と同様に、音声認識部 1 による音声認識結果を対象とした自然言語処理の 1 つである対話管理処理を行うが、その際、音声認識部 1 に対して、必要に応じて変更信号を出力させることが可能である。

【0 1 1 8】

次に、図 1 6 のフローチャートを参照して、音声認識部 1 が図 1 4 に示すように、機械翻訳部 2 が図 1 5 に示すように、それぞれ構成される場合の、図 1 の音声処理システムの動作（翻訳動作）について説明する。

【0 1 1 9】

音声認識部 1 に対して、音声が入力されると、音声認識部 1 では、ステップ S 4 1 において、その音声が音声認識され、その音声認識結果としてのテキストが、機械翻訳部 2 に出力されて、ステップ S 4 2 に進む。

【0 1 2 0】

ステップ S 4 2 では、機械翻訳部 2 において、音声認識部 1 からのテキストを機械翻訳する翻訳処理が行われ、ステップ S 4 3 に進み、音声認識部 1 において、機械翻訳部 2 から、フィードバック情報としての変更信号を受信したかどうかを判定される。

【0 1 2 1】

ステップ S 4 3 において、機械翻訳部 2 から変更信号を受信したと判定された場合、ステップ S 4 1 に戻り、音声認識部 1 は、その変更信号に基づき、音声認識結果の構成単位が変更されるように、再度、音声認識を行い、その音声認識結果を、機械翻訳部 2 に出力する。そして、以下、同様の処理が繰り返される。

【0 1 2 2】

また、ステップ S 4 3 において、機械翻訳部 2 から、フィードバック情報としての変更信号を受信していないと判定された場合、機械翻訳部 2 は、ステップ S 4 2 におけ翻訳処理の結果得られるテキストを、音声合成部 3 に出力し、ステップ S 4 4 に進む。

【0 1 2 3】

ステップ S 4 4 では、音声合成部 3 において、機械翻訳部 2 からのテキストに対応する合成音が生産されて出力され、処理を終了する。

【0 1 2 4】

次に、図 1 7 のフローチャートを参照して、図 1 4 の音声認識部 1 の動作について説明する。

【0 1 2 5】

音声認識部 1 は、音声認識すべき音声が入力されると、ステップ S 5 1 において、その入力音声の音声認識結果の構成単位を設定し、ステップ S 5 2 に進む。ここで、新たな音声が入力された直後においては、ステップ S 5 1 では、所定のデフォルトの構成単位が設定される。

【0 1 2 6】

ステップ S 5 2 では、音声認識部 1 において、入力音声の音声認識が行われ、ステップ S 5 3 に進み、直前のステップ S 5 1 で設定された構成単位の音声認識結果が、機械翻訳部 2 に出力される。そして、ステップ S 5 4 に進み、機械翻訳部 2 から、フィードバック情報としての変更信号を受信したかどうか判定される。ステップ S 5 4 において、変更信号を受信したと判定された場合、ステップ S 5 1 に戻り、前回出力した音声認識結果の構成単位が、変更信号に基づいて小さくまたは大きく設定され、即ち、具体的には、例えば、文節単位から単語単位に設定され、または逆に単語単位から文節単位に設定され、ステップ S 5 2 に進

み、以下、同様の処理が繰り返される。従って、この場合、音声認識部 1 では、変更信号に基づき、構成単位が小さくまたは大きくされた音声認識結果が、その後に行われるステップ S 5 3 で出力される。

【0 1 2 7】

一方、ステップ S 5 4 において、フィードバック情報としての変更信号を受信していないと判定された場合、処理を終了する。

【0 1 2 8】

次に、図 1 8 のフローチャートを参照して、図 1 5 の機械翻訳部 2 の動作について説明する。

【0 1 2 9】

機械翻訳部 2 は、音声認識部 1 から音声認識結果としてのテキストを受信すると、ステップ S 6 1 において、そのテキストの構成単位を解析する。そして、ステップ S 6 2 に進み、その構成単位が、機械翻訳部 2 で処理を行うのに適切な単位であるかどうか判定される。

【0 1 3 0】

ここで、音声認識部 1 からの音声認識結果の構成単位が、機械翻訳部 2 で処理を行うのに適切な単位であるかどうかは、例えば、音声認識結果を形態素解析することにより判定することができる。また、その判定は、例えば、機械翻訳部 2 で処理を行うのに適切な単位の文字列を記憶しておき、その文字列と、音声認識結果の構成単位とを比較すること等によって行うことも可能である。

【0 1 3 1】

ステップ S 6 2 において、音声認識部 1 からの音声認識結果としてのテキストの構成単位が、機械翻訳部 2 で処理を行うのに適切な単位でないと判定された場合、ステップ S 6 3 に進み、構成単位が適切となるように、即ち、構成単位が大きければ小さくし、小さければ大きくするように指示する変更信号が、フィードバック情報として、音声認識部 1 に出力される。そして、音声認識部 1 から、その変更信号にしたがって構成単位が変更された音声認識結果が供給されるのを待って、ステップ S 6 1 に戻り、以下、同様の処理が繰り返される。

【0 1 3 2】

一方、ステップ S 6 2 において、音声認識部 1 からの音声認識結果としてのテキストの構成単位が、機械翻訳部 2 で処理を行うのに適切な単位であると判定された場合、ステップ S 6 4 に進み、機械翻訳部 2 において、音声認識結果が処理される。

【0 1 3 3】

即ち、機械翻訳部 2 では、音声認識結果の翻訳が行われ、その翻訳結果が出力されて、処理を終了する。

【0 1 3 4】

以上のように、音声認識部 1 において、その後段で自然言語処理を行う機械翻訳部 2 からの指示に基づいて、出力する音声認識結果の構成単位を、自然言語処理を行うのに適切な単位に変更して出力するようにしたので、機械翻訳部 2 では、容易に、精度の高い自然言語処理（翻訳）を行うことが可能となる。

【0 1 3 5】

なお、対話管理部 5 においても、上述のような変更信号を、フォーマット情報として、音声認識部 1 に出力し、音声認識部 1 に、その音声認識結果を、対話管理部 5 での処理を行うのに適切な単位で出力させるようにすることが可能である。

【0 1 3 6】

次に、図 1 9 は、図 1 の音声認識部 1 の第 3 の構成例を示している。なお、図中、図 3 における場合と対応する部分については、同一の符号を付してある。即ち、図 1 9 の音声認識部 1 は、適応処理部 1 9 が設けられていないことを除けば、図 3 における場合と基本的に同様に構成されている。

【0 1 3 7】

但し、図 1 9 の実施の形態においては、マッチング部 1 5 は、機械翻訳部 2 から、後述する要求信号を、フィードバック情報として受信することができるようになっており、要求信号を受信した場合、その要求信号による要求に応じた処理を行うようになっている。この場合、マッチング部 1 5 は、一度処理した特徴パラメータが再び必要なときには、特徴量バッファ 1 4 に記憶された特徴パラメータを用いて処理を行うようになっており、これにより、ユーザに、再度発話を

してもらわなくても済むようになっている。

【 0 1 3 8 】

次に、図 2 0 は、音声認識部 1 が図 1 9 に示すように構成される場合の、図 1 の機械翻訳部 2 の構成例（第 3 の構成例）を示している。なお、図中、図 4 における場合と対応する部分については、同一の符号を付してある。即ち、図 2 0 の機械翻訳部 2 は、変換結果修正部 2 9 および変換結果バッファ 3 0 が新たに設けられている他は、図 4 における場合と基本的に同様に構成されている。

【 0 1 3 9 】

但し、図 2 0 の実施の形態においては、言語変換部 2 2 は、その処理の途中結果に基づいて、その処理を精度良く行うために必要な情報があるとき、その情報を要求する要求信号を、フィードバック情報として、図 1 9 の音声認識部 1 （のマッチング部 1 5）に出力し、その情報を得て、精度の良い処理を行うようになっている。

【 0 1 4 0 】

変換結果修正部 2 9 は、言語変換部 2 2 で得られた出力言語の言語情報を受信し、その評価を行う。そして、変換結果修正部 2 9 は、その評価結果が所定の基準を満たしている場合には、言語変換部 2 2 からの言語情報を、そのままテキスト生成部 2 3 に供給する。

【 0 1 4 1 】

一方、言語変換部 2 2 からの言語情報の評価結果が所定の基準を満たしていない場合には、変換結果修正部 2 9 は、その言語情報を、変換結果バッファ 3 0 に一時記憶させるとともに、音声認識結果の他の候補を要求する要求信号を、フィードバック情報として、図 1 9 の音声認識部 1 （のマッチング部 1 5）に出力する。そして、音声認識部 1 が、変換結果修正部 2 9 からの要求信号に応じて、音声認識結果の他の候補を出力し、それが、テキスト解析部 2 1 および言語変換部 2 2 で処理され、出力言語の言語情報とされて、変換結果修正部 2 9 に供給されると、変換結果修正部 2 9 は、その新たに受信した言語情報と、変換バッファ 3 0 に記憶された言語情報とを比較し、入力テキストの翻訳結果として、より適切な方を選択して、テキスト生成部 2 3 に出力する。

【0142】

次に、図21は、図20の言語変換部22および言語変換データベース26の構成例を示している。

【0143】

言語変換部22は、マッチング部51から構成されており、マッチング部51は、言語変換データベース26を参照することで、テキスト解析部21から供給される入力テキストの言語情報を、出力言語の言語情報に変換する。

【0144】

即ち、図21の実施の形態では、言語変換データベース26は、シソーラス辞書52およびテンプレートテーブル53から構成されている。シソーラス辞書52は、例えば、図22に示すように、単語およびその概念（意味）が階層的に分類されたシソーラスを記憶している。なお、図22においては、長方形で囲んである記載が概念を表し、楕円形で囲んである記載が単語を表している。また、概念を表す長方形に記載されている番号は、シソーラスにおいて、その概念を特定するための分類番号である。

【0145】

図21に戻り、テンプレートテーブル53には、日本語の文のパターンと、その英語訳とを対応付けたテンプレートが登録されている。なお、テンプレートでは、日本語文のパターンは、必要に応じて変数（図21においては、XやY）を用いて記述されている。また、変数には番号が付されており、この番号は、図22に示したシソーラス上の分類番号を表す。

【0146】

以上のように構成される言語変換部22では、マッチング部51において、テンプレートテーブル53に記憶されているテンプレートから、入力テキストと最も類似する（距離の近い）日本語文のパターンを有するものが選択される。即ち、マッチング部51は、テンプレートテーブル53の各テンプレートの日本語文のパターンと、入力テキストとの距離を求め、その距離を最も短くするテンプレートを選択する。さらに、その選択されたテンプレート（以下、適宜、選択テンプレートという）の日本語文のパターンの中の変数に対応する語句（以下、適宜

、対応語句という)が、入力テキストから抽出され、その対応語句と、選択テンプレートの変数が属する概念との間の意味的な距離(以下、適宜、単語間距離という)が求められる。

【0147】

ここで、選択テンプレートの変数と、その対応語句との単語距離としては、例えば、シソーラスにおいて、その変数のノードから、対応語句のノードに移動するための最小移動数(変数のノードから、対応語句のノードへの最短のパスを構成するブランチの数)などを用いることができる。なお、選択テンプレートに、複数の変数が存在する場合には、その複数の変数それぞれについて、単語間距離が計算される。

【0148】

マッチング部51は、選択テンプレートを決定し、さらに、その選択テンプレートにおける変数についての単語間距離を求めると、選択テンプレート、対応語句、および単語間距離とともに、選択テンプレートにおける日本語文のパターンと入力テキストとの距離(これは、上述のように、選択テンプレートを決定するときに求められる)(以下、適宜、パターン間距離という)を、変換結果修正部29に出力する。

【0149】

具体的には、例えば、入力テキストが「金槌を使った」であった場合において、日本語文「X(1.5621)を使う」を有するテンプレートが、選択テンプレートとして決定(選択)されたとすると、その選択テンプレート、対応語句「金槌」、入力テキスト「金槌を使った」と日本語文「X(1.5621)を使う」とのパターン間距離、並びに変数X(1.5621)とその対応語句「金槌」との単語間距離が、変換結果修正部29に出力される。

【0150】

なお、マッチング部51は、基本的には、上述のように、入力テキストとのパターン間距離を最も短くするテンプレートを、そのまま、選択テンプレートとして決定するが、後述するように、2つなどの複数のテンプレートが、選択テンプレートの候補として得られ、最終的な選択テンプレートを決定することが困難な

場合には、音声認識部 1 に対して、より確からしいテンプレートを、選択テンプレートとして決定するのに必要な情報を要求する要求信号を出力し、その要求した情報が、音声認識部 1 から供給されるのを待って、選択テンプレートを決定するようになっている。

【0151】

次に、図 23 のフローチャートを参照して、図 20 の機械翻訳部 2 の動作について、さらに説明する。

【0152】

機械翻訳部 2 では、図 19 の音声認識部 1 から音声認識結果としての入力テキストを受信すると、ステップ S71 において、変換結果バッファ 30 の記憶内容がクリアされ、ステップ S72 に進み、テキスト解析部 21 は、入力テキストの解析を行う。テキスト解析部 21 による解析結果は、言語変換部 22 に供給され、言語変換部 22 は、ステップ S73 において、図 21 で説明したような選択テンプレートの決定等を行い、その選択テンプレートを用いて、入力テキストの言語情報を出力テキストの言語情報に変換し、その変換結果としての、上述した選択テンプレート、パターン間距離、対応語句、および単語間距離を、変換結果修正部 29 に出力する。

【0153】

変換結果修正部 29 は、ステップ S74 において、言語変換部 22 からの出力テキストの言語情報（選択テンプレート、パターン間距離、対応語句および単語間距離）を、変換結果バッファ 30 に供給して記憶させ、ステップ S75 に進み、言語変換部 22 から供給された単語間距離が、所定の基準値以下であるかどうかを判定する。ステップ S75 において、単語間距離が、所定の基準値以下であると判定された場合、即ち、選択テンプレートの変数が属する概念と、その変数についての入力テキストの対応語句との意味的な距離が近く、従って、入力テキストの対応語句が、正しい音声認識結果であると推測される場合、ステップ S76 に進み、変換結果修正部 29 は、ステップ S74 で変換結果バッファ 30 に記憶させた出力言語の言語情報を、テキスト生成部 23 に出力する。そして、テキスト生成部 23 では、入力テキストを出力言語に翻訳したテキストが生成され、

処理を終了する。

【0154】

一方、ステップS75において、単語間距離が、所定の基準値以下でないと判定された場合、即ち、選択テンプレートの変数が属する概念と、その変数についての入力テキストの対応語句との意味的な距離が遠く、従って、入力テキストの対応語句が、誤った音声認識結果であると推測される場合（ここでは、音響的には正しいが、意味的に誤っていると推測される場合）、ステップS77に進み、変換結果修正部29は、対応語句について、例えば、その対応語句と同音異義語の他の候補を要求する要求信号を、図19の音声認識部1に出力する。

【0155】

この場合、図19の音声認識部1では、特徴量バッファ14に記憶された特徴パラメータを用いて音声認識が再度行われ、これにより、対応語句と同音異義語の他の候補が、図20の機械翻訳部2に供給される。なお、対応語句と同音異義語を対象とした音声認識は、例えば、図19の音声認識部1の辞書データベース17に、各種の同音異義語を記憶しておくことで行うことが可能である。

【0156】

対応語句と同音異義語の他の候補が、図19の音声認識部1から図20の機械翻訳部2に供給されると、ステップS78において、テキスト解析部21および言語変換部22は、入力テキストの対応語句を、その同音異義語の他の候補に置き換えたもの（以下、適宜、置換テキストという）を対象に処理を行う。そして、その処理結果は、変換結果修正部29に出力される。

【0157】

なお、対応語句と同音異義語の他の候補が、複数存在する場合には、音声認識部1から機械翻訳部2に対しては、その複数の候補が供給される。この場合、機械翻訳部2では、その複数の候補それぞれについて、置換テキストが作成されて処理される。

【0158】

以上のようにして、変換結果修正部29は、言語変換部22から、置換テキストを変換した出力言語の言語情報を受信すると、ステップS79において、その

言語情報と、変換結果バッファ 30 に記憶されている言語情報のうち、より適切なものを選択する。即ち、変換結果修正部 29 は、置換テキストを変換した出力言語の言語情報と、変換結果バッファ 30 に記憶されている言語情報のうち、その言語情報に含まれる単語間距離が最も小さいもの（選択テンプレートの変数が属する概念との意味的な距離が近い対応語句を有するテキストを変換した言語情報）を選択する。

【0159】

そして、ステップ S 76 に進み、変換結果修正部 29 は、選択した言語情報を、テキスト生成部 23 に出力し、以下、上述した場合と同様の処理が行われ、処理を終了する。

【0160】

なお、置換テキストが複数存在する場合には、ステップ S 79 では、その複数の置換テキストそれぞれを変換した言語情報と、変換結果バッファ 30 に記憶されている言語情報の中から、単語間距離が最も短い言語情報が選択される。

【0161】

以上の処理によれば、例えば、入力テキストが「蜘蛛が白い」であり、選択テンプレートとして、日本語文「X (1. 4829) が白い」を有するテンプレートが決定された場合、変数 X の対応語句は「蜘蛛」となるが、この場合、分類番号が 1. 4829 である概念と、その対応語句「蜘蛛」との意味的な距離が遠いときには、変換結果修正部 29 において、上述したような要求信号が、フィードバック情報として、図 19 の音声認識部 1 に出力される。そして、いま、音声認識部 1 が、その要求信号に応じて、「くも」の同音異義語である、「蜘蛛」の他の候補として「雲」を、機械翻訳部 2 に供給したとすると、機械翻訳部 2 では、「蜘蛛」と「雲」のうち、分類番号が 1. 4829 である概念との意味的な距離が近い方が、最終的な対応語句とされる。

【0162】

従って、音声認識部 1 において、音響的には正しいが、意味的に誤った音声認識結果が得られた場合であっても、その意味的に誤った音声認識結果を修正し、これにより、正しい翻訳結果を得ることが可能となる。

【0163】

次に、図24のフローチャートを参照して、図21のマッチング部51で行われる、選択テンプレートを決定する処理（テンプレートテーブル53からテンプレートを選択する処理）について説明する。

【0164】

ステップS81では、テンプレートテーブル53から、あるテンプレートが注目すべき注目テンプレートとして選択され、ステップS82に進み、その注目テンプレート（に記述された日本語パターン）と、入力テキストとのパターン間距離が計算される。そして、ステップS83に進み、テンプレートテーブル53に記憶されたすべてのテンプレートについて、パターン間距離を計算したかどうか判定される。ステップS83において、テンプレートテーブル53に記憶されたすべてのテンプレートについて、まだ、パターン間距離を計算していないと判定された場合、ステップS81に戻り、まだ注目テンプレートとされていないテンプレートが、新たに注目テンプレートとして選択され、以下、同様の処理が繰り返される。

【0165】

また、ステップS83において、テンプレートテーブル53に記憶されたすべてのテンプレートについて、入力テキストとのパターン間距離を計算したと判定された場合、ステップS84に進み、パターン間距離が最も小さいテンプレート（以下、適宜、第1位のテンプレートという）と、パターン間距離が2番目に小さいテンプレート（以下、適宜、第2位のテンプレートという）とが検出され、その第1位および第2位のテンプレートについてのパターン間距離の差が、所定の閾値以下であるかどうか判定される。

【0166】

ステップS84において、パターン間距離の差が、所定の閾値以下でないと判定された場合、即ち、第1位のテンプレートに記述された日本語文が、テンプレートテーブル53に記憶された他のテンプレートのいずれに記述された日本語文と比較しても、入力テキストに対して、非常に良く合致する場合、ステップS85に進み、その第1位のテンプレートが、選択テンプレートとして決定され、処

理を終了する。

【0167】

一方、ステップS84において、パターン間距離の差が、所定の閾値以下であると判定された場合、即ち、入力テキストが、第1位のテンプレートに記述された日本語文だけでなく、第2位のテンプレートに記述された日本語文とも比較的合致する場合、ステップS86に進み、マッチング部51は、その第1位と第2位のテンプレートから、より入力音声に合致する日本語文を有する方を決定するための音響的な評価値を要求する要求信号を、フィードバック情報として、図19の音声認識部1に供給する。

【0168】

この場合、音声認識部1では、特徴量バッファ14に記憶された特徴パラメータを用い、入力音声の音声認識結果を、第1位のテンプレートに記述された日本語文とすることの尤度と、第2位のテンプレートに記述された日本語文とすることの尤度が求められ、図20の機械翻訳部2に供給される。

【0169】

機械翻訳部2では、図19の音声認識部1からの第1位と第2位のテンプレートに記述された日本語文それぞれについての尤度が、テキスト解析部21を介して、言語変換部22のマッチング部51に供給され、マッチング部51は、このようにして第1位と第2位のテンプレートに記述された日本語文それぞれについての尤度を受信すると、ステップS87において、その尤度の高い方のテンプレートを、選択テンプレートとして決定し、処理を終了する。

【0170】

以上の処理によれば、例えば、入力音声の音声認識結果として「金槌を使い」が得られ、その音声認識結果「金槌を使い」に対して、日本語文「X(1, 23)を使う」を有するテンプレートと、日本語文「X(1, 23)を使え」を有するテンプレートが、それぞれ第1位と第2位のテンプレートであり、音声認識結果「金槌を使い」と第1位のテンプレートとのパターン間距離と、音声認識結果「金槌を使い」と第2位のテンプレートとのパターン間距離とが近い値である場合には、音声認識部1において、入力音声の音声認識結果を、「金槌を使う」と

することの尤度と、「金槌を使え」とすることの尤度が計算される。そして、機械翻訳部 2 では、その尤度の高い方の日本語文を有するテンプレートが、選択テンプレートとして決定される。

【0171】

従って、音声認識部 1 において、誤った音声認識結果が得られた場合であっても、その誤った音声認識結果を修正し、これにより、正しい翻訳結果を得ることが可能となる。

【0172】

なお、図 24 のフローチャートにしたがった処理は、第 3 位以下のテンプレートも対象にして行うことが可能である。

【0173】

次に、図 23 のフローチャートにしたがった処理においては、同音異義語の中から、選択テンプレートにおける変数が属する概念との意味的な距離が最も近い単語を選択することで、誤った音声認識結果を修正するようにしたが、この処理では、同音異義語の中に、選択テンプレートにおける変数が属する概念との意味的な距離が近いものが複数存在する場合に対処するのが困難である。

【0174】

即ち、例えば、選択テンプレート「X (1. 4 8 3 0) で食べた」の X (1. 4 8 3 0) に対応する同音異義語の対応語句として、「橋」、「端」、および「箸」の 3 つが得られた場合に、その「橋」、「端」、「箸」との意味的な距離がいずれも同じ値であるときには、その優劣をつけることができない。

【0175】

そこで、このような場合には、図 20 の機械翻訳部 2 において、それらの同音異義語に対応する入力音声の部分の、例えば、アクセントやピッチ等のプロソディ (Prosody) に基づいて、音声認識結果として最も確からしい単語を決定するように要求する要求信号を、フィードバック情報として、図 19 の音声認識部 1 に供給するようにすることができる。

【0176】

即ち、例えば、上述の「橋」、「端」、「箸」については、それぞれの発話に

において、図 2 5 に示すような抑揚があるのが一般的であり、従って、音声認識部 1 では、特徴量バッファ 1 4 に記憶された特徴パラメータに基づき、入力音声のプロソディを得て、そのプロソディが、「橋」、「端」、「箸」のうちのいずれに最も合致するかを検出することで、音声認識結果として最も確からしい単語を決定することができる。

【0 1 7 7】

そこで、図 2 6 のフローチャートを参照して、上述のような要求信号を出力する場合の、図 2 0 の機械翻訳部 2 の動作について説明する。

【0 1 7 8】

図 2 0 の機械翻訳部 2 では、ステップ S 9 1 乃至 S 9 8 において、図 2 3 のステップ S 7 1 乃至 S 7 8 における場合とそれぞれ同様の処理が行われる。

【0 1 7 9】

そして、ステップ S 9 8 の処理後は、ステップ S 9 9 に進み、変換結果修正部 2 9 は、言語変換部 2 2 から供給される、置換テキストを変換した出力言語の言語情報と、変換結果バッファ 3 0 に記憶されている言語情報に含まれる単語間距離が同一であるかどうかを判定する。ステップ S 9 9 において、単語間距離が同一でないと判定された場合、ステップ S 1 0 0 に進み、変換結果修正部 2 9 は、図 2 3 のステップ S 9 における場合と同様に、置換テキストを変換した出力言語の言語情報と、変換結果バッファ 3 0 に記憶されている言語情報のうち、その言語情報に含まれる単語間距離が最も小さいものを選択する。

【0 1 8 0】

そして、ステップ S 9 6 に進み、変換結果修正部 2 9 は、選択した言語情報を、テキスト生成部 2 3 に出力し、これにより、テキスト生成部 2 3 では、入力テキストを出力言語に翻訳したテキストが生成され、処理を終了する。

【0 1 8 1】

一方、ステップ S 9 9 において、言語変換部 2 2 から供給される、置換テキストを変換した出力言語の言語情報と、変換結果バッファ 3 0 に記憶されている言語情報（入力テキストを変換した出力言語の言語情報）に含まれる単語間距離が同一であると判定された場合、ステップ S 1 0 1 に進み、変換結果修正部 2 9 は

、置換テキストと入力テキストに含まれる同音異義語に対応する入力音声の部分のプロソディに基づいて、音声認識結果として最も確からしい単語を決定するように要求する要求信号を、フィードバック情報として、図 19 の音声認識部 1 に供給する。

【0182】

この場合、音声認識部 1 は、変換結果修正部 29 からの要求信号に応じて、上述したように、入力音声のプロソディに基づき、同音異義語の中から、音声認識結果として最も確からしい単語（以下、適宜、最尤単語という）を決定し、図 20 の機械翻訳部 2 に供給する。

【0183】

この最尤単語は、テキスト解析部 21 および言語変換部 22 を介して、変換結果修正部 29 に供給され、変換結果修正部 29 は、ステップ S102 において、音声認識部 1 からの最尤単語を有する言語情報を選択し、ステップ S96 に進む。ステップ S96 では、変換結果修正部 29 は、選択した言語情報を、テキスト生成部 23 に出力し、これにより、テキスト生成部 23 では、入力テキストを出力言語に翻訳したテキストが生成され、処理を終了する。

【0184】

次に、図 27 は、図 1 の音声認識部 1 の第 4 の構成例を示している。なお、図中、図 3 における場合と対応する部分については、同一の符号を付してある。即ち、図 27 の音声認識部 1 は、適応処理部 19 が除去され、かつ分野別辞書群 20 が新たに設けられている他は、図 3 における場合と同様に構成されている。

【0185】

分野別辞書群 20 は、N 個の分野別辞書から構成されており、各分野別辞書は、基本的には、辞書データベース 17 の単語辞書と同様に構成されている。但し、各分野別辞書は、特定の話題や分野等に特化した単語（語句）に関する言語モデル、即ち、タスク別の言語モデルを記憶している。

【0186】

そして、図 27 の実施の形態では、マッチング部 15 は、通常は、音響モデルデータベース 16、辞書データベース 17、および文法データベース 18 だけを

参照して処理を行うが、機械翻訳部 2 からの要求信号に応じて、分野別辞書群 2 0 の必要な分野別辞書をさらに参照して処理を行うようになっている。

【0 1 8 7】

次に、図 2 8 は、音声認識部 1 が図 2 7 に示したように構成される場合の、図 1 の機械翻訳部 2 の第 4 の構成例を示している。なお、図中、図 2 0 における場合と対応する部分については、同一の符号を付してある。即ち、図 2 8 の機械翻訳部 2 は、分野推定部 6 1 が新たに設けられている他は、図 2 0 における場合と同様に構成されている。

【0 1 8 8】

分野推定部 6 1 には、テキスト解析部 2 1 から、入力テキストの解析結果が供給されるとともに、変換結果修正部 2 9 が出力する要求信号が供給されるようになっている。そして、分野推定部 6 1 は、テキスト解析部 2 1 からの解析結果に基づいて、即ち、それまでに入力された音声の音声認識結果の解析結果に基づいて、ユーザの発話内容の分野や話題等のタスクを推定し、変換結果修正部 2 9 から要求信号を受信した場合には、推定した分野や話題に対応する分野別辞書を追加して処理を行うように要求する要求信号を、フィードバック情報として、図 2 7 の音声認識部 1 に供給するようになっている。

【0 1 8 9】

ここで、発話から、その内容の分野や話題を推定する方法については、例えば、「自然言語検索システムにおける分野推論方式」、別所克人、岩瀬成人、戸部美春、福村好美、電子情報通信学会論文誌 D-II J81-DII No.6 pp.1317-1327 などに、その詳細が記載されている。

【0 1 9 0】

次に、図 2 9 のフローチャートを参照して、図 2 8 の機械翻訳部 2 の動作について説明する。

【0 1 9 1】

図 2 8 の機械翻訳部 2 では、ステップ S 1 1 1 乃至 S 1 1 4 において、図 2 3 のステップ S 7 1 乃至 S 7 4 における場合とそれぞれ同様の処理が行われる。

【0 1 9 2】

そして、ステップ S 1 1 4 の処理後は、ステップ S 1 1 5 に進み、変換結果修正部 2 9 は、言語変換部 2 2 から供給されたパターン間距離が、所定の基準値以下であるかどうかを判定する。ステップ S 1 1 5 において、パターン間距離が、所定の基準値以下であると判定された場合、即ち、音声認識結果と、選択テンプレートに記述された日本語文との距離が近く、従って、音声認識部 1 (図 2 7) において、分野別辞書群 2 0 を用いなくても、正しい音声認識結果が得られていると推測される場合、ステップ S 1 1 6 に進み、変換結果修正部 2 9 は、ステップ S 1 1 4 で変換結果バッファ 3 0 に記憶させた出力言語の言語情報を、テキスト生成部 2 3 に出力する。そして、テキスト生成部 2 3 では、入力テキストを出力言語に翻訳したテキストが生成され、処理を終了する。

【0 1 9 3】

一方、ステップ S 1 1 5 において、パターン間距離が、所定の基準値以下でないと判定された場合、即ち、音声認識結果と、選択テンプレートに記述された日本語文との距離が遠く、従って、音声認識部 1 (図 2 7) において、分野別辞書群 2 0 をさらに用いて処理を行わないと、正しい音声認識結果が得られないと推測される場合、ステップ S 1 1 7 に進み、変換結果修正部 2 9 は、分野別辞書群 2 0 を用いた音声認識を要求する要求信号を、分野推定部 6 1 に供給する。

【0 1 9 4】

分野推定部 6 1 は、テキスト解析部 2 1 の出力を参照することで、発話内容の話題や分野を推定しており、変換結果修正部 2 9 から要求信号を受信すると、推定した分野や話題に対応する分野別辞書を追加して処理を行うように要求する要求信号を、フィードバック情報として、図 2 7 の音声認識部 1 に供給する。

【0 1 9 5】

即ち、分野推定部 6 1 は、例えば、発話内容の話題が旅行に関することであると推定した場合には、観光地の地名が登録されている分野別辞書を追加し、これにより、音声認識の対象語彙として、観光地の地名を含めて処理を行うように要求する要求信号を、図 2 7 の音声認識部 1 に供給する。

【0 1 9 6】

この場合、音声認識部 1 では、特徴量バッファ 1 4 に記憶された特徴パラメー

タを用い、要求信号に応じた分野や話題に関する単語（語句）の情報が登録されている分野別辞書をさらに参照して音声認識が行われ、これにより、音声認識の対象語彙を、いわば拡張した音声認識が行われる。そして、その音声認識結果は、音声認識部 1 から機械翻訳部 2 に供給される。

【0197】

上述のようにして、新たな音声認識結果が、図 2 7 の音声認識部 1 から供給されると、図 2 8 の機械翻訳部 2 では、ステップ S 1 1 8 において、テキスト解析部 2 1 および言語変換部 2 2 は、その新たな音声認識結果としての入力テキストを対象に処理を行う。そして、その処理結果は、変換結果修正部 2 9 に出力される。

【0198】

以上のようにして、変換結果修正部 2 9 は、言語変換部 2 2 から、新たな音声認識結果を変換した出力言語の言語情報を受信すると、ステップ S 1 1 9 において、その言語情報と、変換結果バッファ 3 0 に記憶されている言語情報のうち、より適切なものを選択する。即ち、変換結果修正部 2 9 は、新たな音声認識結果を変換した出力言語の言語情報と、変換結果バッファ 3 0 に記憶されている言語情報のうち、例えば、パターン間距離が小さいものを選択する。

【0199】

そして、ステップ S 1 1 6 に進み、変換結果修正部 2 9 は、選択した言語情報を、テキスト生成部 2 3 に出力し、以下、上述した場合と同様の処理が行われ、処理を終了する。

【0200】

以上のように、機械翻訳部 2 において、その処理の途中結果に基づいて、音声認識部 1 に対して、要求信号を、フィードバック情報としてフィードバックし、音声認識部 1 において、その要求に応じた処理を行うようにした場合には、音声に対して、容易に、高度な自然言語処理を施すことが可能となる。

【0201】

即ち、この場合、音声認識部 1 では、比較的簡易な音声認識処理を行い、機械翻訳部 2 において、その音声認識結果の処理の最中に、疑問が生じたときや、新

たな情報が必要になったときに、音声認識部 1 にその疑問を解決するための処理や必要な情報を要求することができる。その結果、機械翻訳部 2 では、音声に対して、容易に、高度な自然言語処理を施すことが可能となる。

【0202】

また、この場合、発話者に、再度発話を行うように催促したり、音声認識結果が正しいかどうかを確認する必要もない。

【0203】

なお、本実施の形態では、機械翻訳部 2 において、日本語文のパターンを有するテンプレートを用いて翻訳を行うようにしたが、翻訳は、その他、例えば、用例を用いて行うようにすることも可能である。

【0204】

次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、専用のハードウェアとしての音声処理システムに組み込まれているコンピュータや、汎用のコンピュータ等にインストールされる。

【0205】

そこで、図 30 を参照して、上述した一連の処理を実行するプログラムをコンピュータにインストールし、コンピュータによって実行可能な状態とするために用いられる、そのプログラムが記録されている記録媒体について説明する。

【0206】

プログラムは、図 30 (A) に示すように、コンピュータ 101 に内蔵されている記録媒体としてのハードディスク 102 や半導体メモリ 103 に予め記録しておくことができる。

【0207】

あるいはまた、プログラムは、図 30 (B) に示すように、フロッピーディスク 111、CD-ROM (Compact Disc Read Only Memory) 112、MO (Magneto optical) ディスク 113、DVD (Digital Versatile Disc) 114、磁気ディスク 115、半導体メモリ 116 などの記録媒体に、一時的あるいは永続的に格納（記録）

しておくことができる。このような記録媒体は、いわゆるパッケージソフトウェアとして提供することができる。

【0208】

なお、プログラムは、上述したような記録媒体からコンピュータにインストールする他、図30(C)に示すように、ダウンロードサイト121から、デジタル衛星放送用の人工衛星122を介して、コンピュータ101に無線で転送したり、LAN(Local Area Network)、インターネットといったネットワーク131を介して、コンピュータ123に有線で転送し、コンピュータ101において、内蔵するハードディスク102などにインストールすることができる。

【0209】

ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【0210】

また、プログラムは、1のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【0211】

次に、図31は、図30のコンピュータ101の構成例を示している。

【0212】

コンピュータ101は、図31に示すように、CPU(Central Processing Unit)142を内蔵している。CPU142には、バス141を介して、入出力インタフェース145が接続されており、CPU142は、入出力インタフェース145を介して、ユーザによって、キーボードやマウス等で構成される入力部147が操作されることにより指令が入力されると、それにしたがって、図30(A)の半導体メモリ103に対応するROM(Read Only Memory)143に格納されているプログラムを実行する。あるいは、また、CPU142は、ハードディスク102に

格納されているプログラム、衛星 1 2 2 若しくはネットワーク 1 3 1 から転送され、通信部 1 4 8 で受信されてハードディスク 1 0 2 にインストールされたプログラム、またはドライブ 1 4 9 に装着されたフロッピディスク 1 1 1、CD-ROM 1 1 2、MOディスク 1 1 3、DVD 1 1 4、若しくは磁気ディスク 1 1 5 から読み出されてハードディスク 1 0 2 にインストールされたプログラムを、RAM(Random Access Memory) 1 4 4 にロードして実行する。そして、CPU 1 4 2 は、その処理結果を、例えば、入出力インタフェース 1 4 5 を介して、LCD(Liquid Crystal Display)等で構成される表示部 1 4 6 に、必要に応じて出力する。

【0 2 1 3】

【発明の効果】

本発明の音声処理装置および音声処理方法、並びに記録媒体によれば、音声認識結果を対象とする自然言語処理によって得られる情報がフィードバックされ、そのフィードバックされる情報に基づいて処理が行われる。従って、例えば、音声認識に用いるモデルの適応を、ユーザに負担をかけることなく、かつ精度良く行うこと等が可能となる。

【図面の簡単な説明】

【図 1】

本発明を適用した音声処理システムの一実施の形態の構成例を示すブロック図である。

【図 2】

図 1 の音声処理システムの動作の概要を説明するための図である。

【図 3】

音声認識部 1 の第 1 の構成例を示すブロック図である。

【図 4】

機械翻訳部 2 の第 1 の構成例を示すブロック図である。

【図 5】

音声合成部 3 の構成例を示すブロック図である。

【図 6】

対話管理部 5 の第 1 の構成例を示すブロック図である。

【図 7】

音声処理システムの第 1 の動作例を説明するためのフローチャートである。

【図 8】

対話管理部 5 の第 2 の構成例を示すブロック図である。

【図 9】

音声処理システムの第 2 の動作例を説明するためのフローチャートである。

【図 1 0】

対話管理部 5 の第 3 の構成例を示すブロック図である。

【図 1 1】

音声処理システムの第 3 の動作例を説明するためのフローチャートである。

【図 1 2】

対話管理部 5 の第 4 の構成例を示すブロック図である。

【図 1 3】

音声処理システムの第 4 の動作例を説明するためのフローチャートである。

【図 1 4】

音声認識部 1 の第 2 の構成例を示すブロック図である。

【図 1 5】

機械翻訳部 2 の第 2 の構成例を示すブロック図である。

【図 1 6】

音声処理システムの第 5 の動作例を説明するためのフローチャートである。

【図 1 7】

図 1 4 の音声認識部 1 の動作を説明するためのフローチャートである。

【図 1 8】

図 1 5 の機械翻訳部 2 の動作を説明するためのフローチャートである。

【図 1 9】

音声認識部 1 の第 3 の構成例を示すブロック図である。

【図 2 0】

機械翻訳部 2 の第 3 の構成例を示すブロック図である。

【図 2 1】

図 2 0 の言語処理部 2 2 の構成例を示すブロック図である。

【図 2 2】

シソーラスを示す図である。

【図 2 3】

図 2 0 の機械翻訳部 2 の第 1 の動作例を説明するためのフローチャートである。

【図 2 4】

マッチング部 5 1 において行われるテンプレートを選択する処理を説明するためのフローチャートである。

【図 2 5】

「橋」、「端」、および「箸」のアクセントを示す図である。

【図 2 6】

図 2 0 の機械翻訳部 2 の第 2 の動作例を説明するためのフローチャートである。

【図 2 7】

音声認識部 1 の第 4 の構成例を示すブロック図である。

【図 2 8】

機械翻訳部 2 の第 4 の構成例を示すブロック図である。

【図 2 9】

図 2 8 の機械翻訳部 2 の動作を説明するためのフローチャートである。

【図 3 0】

本発明を適用した記録媒体を説明するための図である。

【図 3 1】

図 3 0 のコンピュータ 1 0 1 の構成例を示すブロック図である。

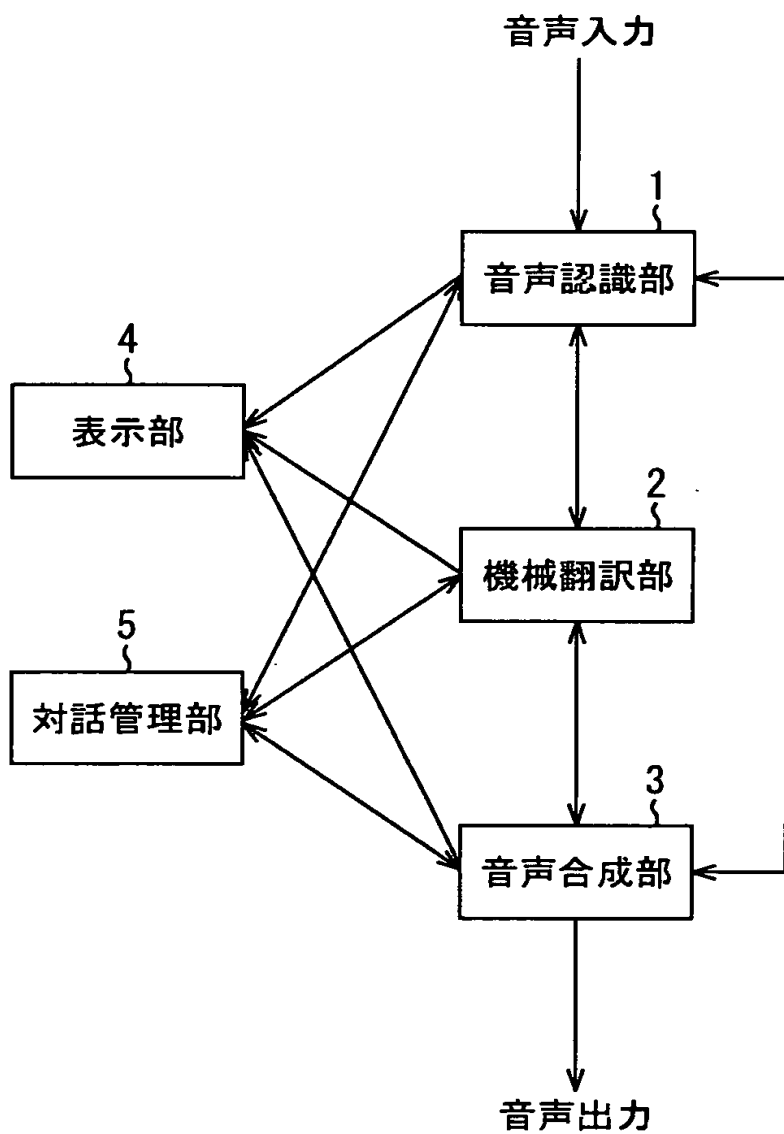
【符号の説明】

1 音声認識部, 2 機械翻訳部, 3 音声合成部, 4 表示部, 5
対話管理部, 1 1 マイク (マイクロフォン), 1 2 AD変換部, 1
3 特徴抽出部, 1 4 特徴量バッファ, 1 5 マッチング部, 1 6 音
響モデルデータベース, 1 7 辞書データベース, 1 8 文法データベース

, 1 9 適応処理部, 2 0 分野別辞書群, 2 1 テキスト解析部, 2
2 言語変換部, 2 3 テキスト生成部, 2 4 辞書データベース, 2 5
解析用文法データベース, 2 6 言語変換データベース, 2 7 辞書デー
タベース, 2 8 生成用文法データベース, 2 9 変換結果修正部, 3 0
変換結果バッファ, 3 1 テキスト解析部, 3 2 規則合成部, 3 3
DA変換部, 3 4 辞書データベース, 3 5 解析用文法データベース,
3 6 音素片データベース, 4 1 対話処理部, 4 2 音声区間検出部,
4 3 修正部, 4 4 信頼度計算部, 4 5 タスク推定部, 5 1 マッチ
ング部, 5 2 シソーラス辞書, 5 3 テンプレートテーブル, 6 1 分
野推定部, 1 0 1 コンピュータ, 1 0 2 ハードディスク, 1 0 3 半
導体メモリ, 1 1 1 フロッピーディスク, 1 1 2 CD-ROM, 1 1 3 MO
ディスク, 1 1 4 DVD, 1 1 5 磁気ディスク, 1 1 6 半導体メモリ
, 1 2 1 ダウンロードサイト, 1 2 2 衛星, 1 3 1 ネットワーク,
1 4 1 バス, 1 4 2 CPU, 1 4 3 ROM, 1 4 4 RAM, 1 4 5
入出カインタフェース, 1 4 6 表示部, 1 4 7 入力部, 1 4 8 通信
部, 1 4 9 ドライブ

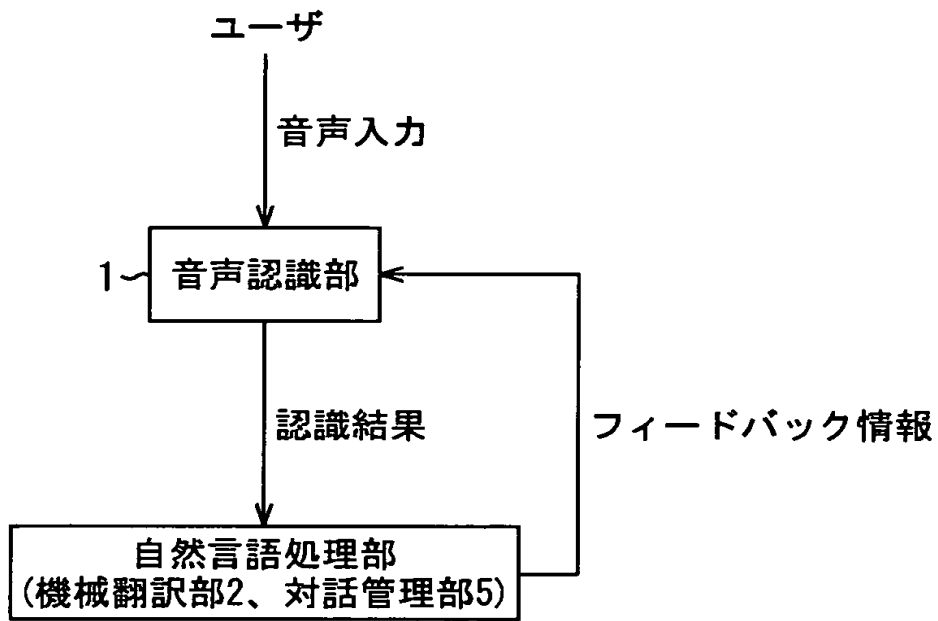
【書類名】 図面

【図 1】

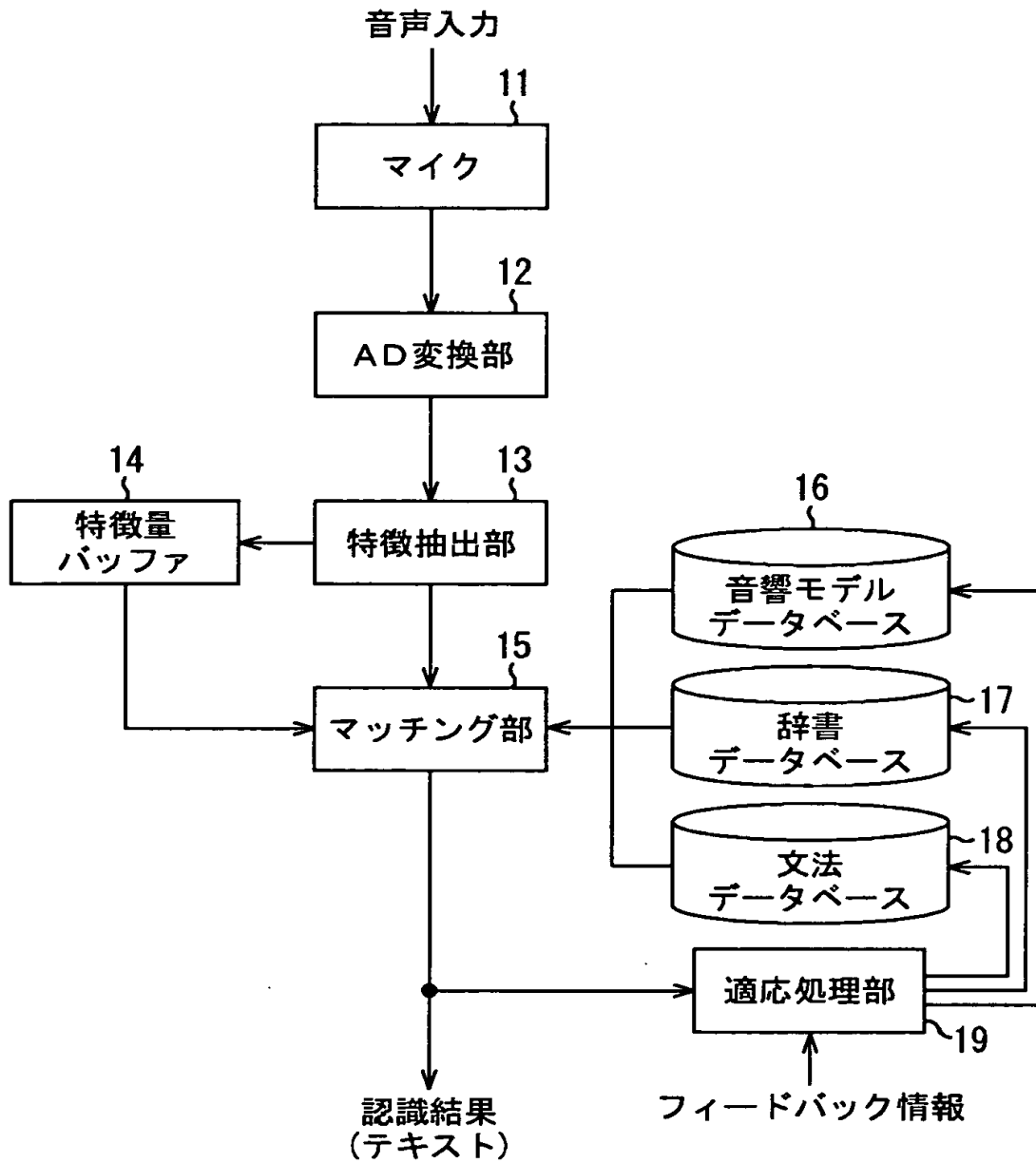


音声処理システム

【図 2】

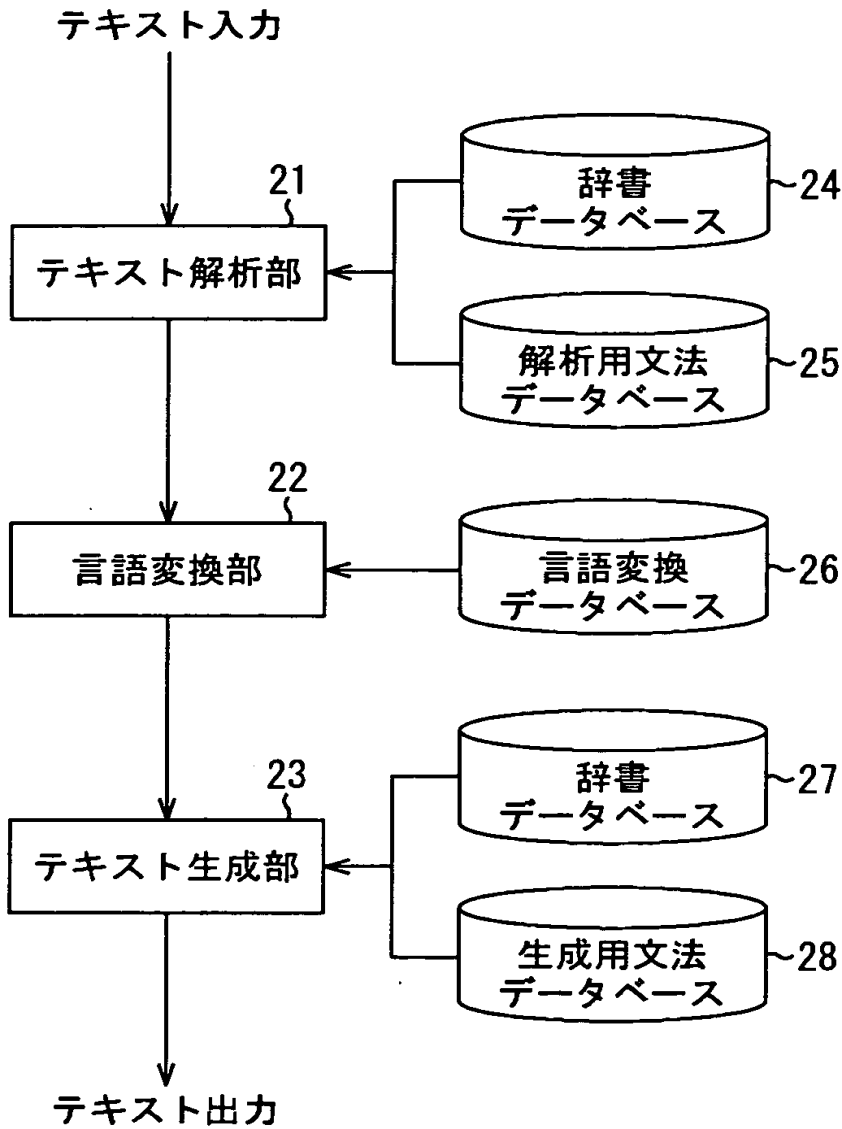


【図 3】



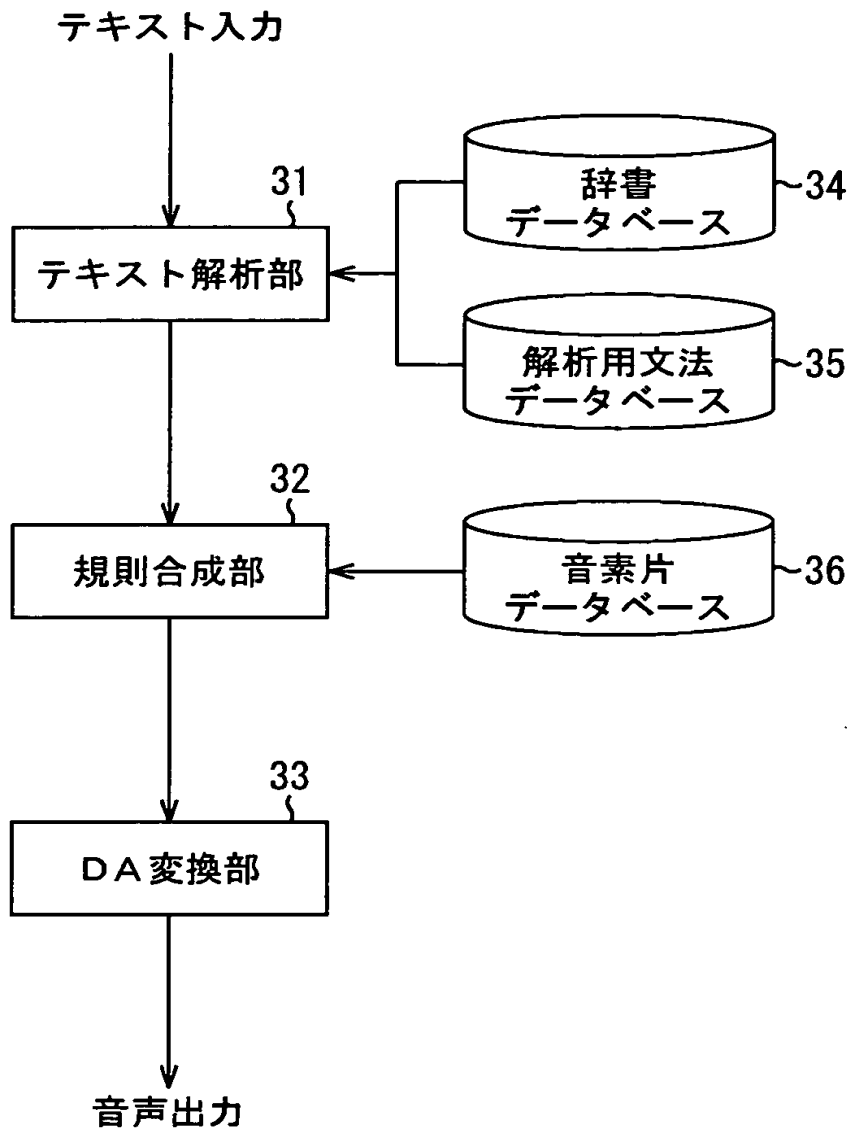
音声認識部 1

【図 4】



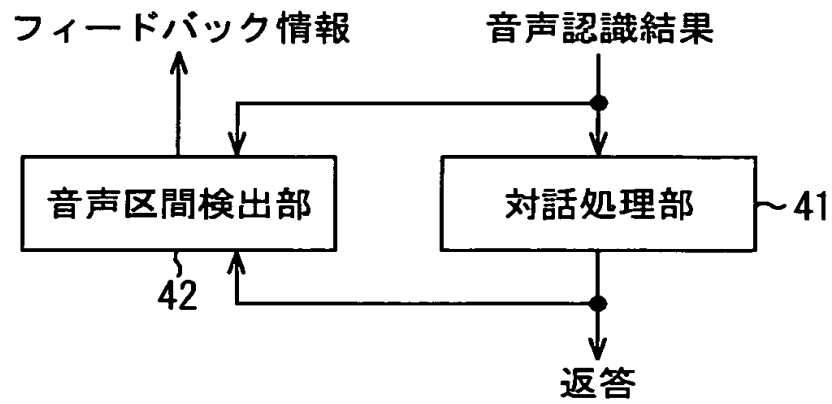
機械翻訳部 2

【図 5】



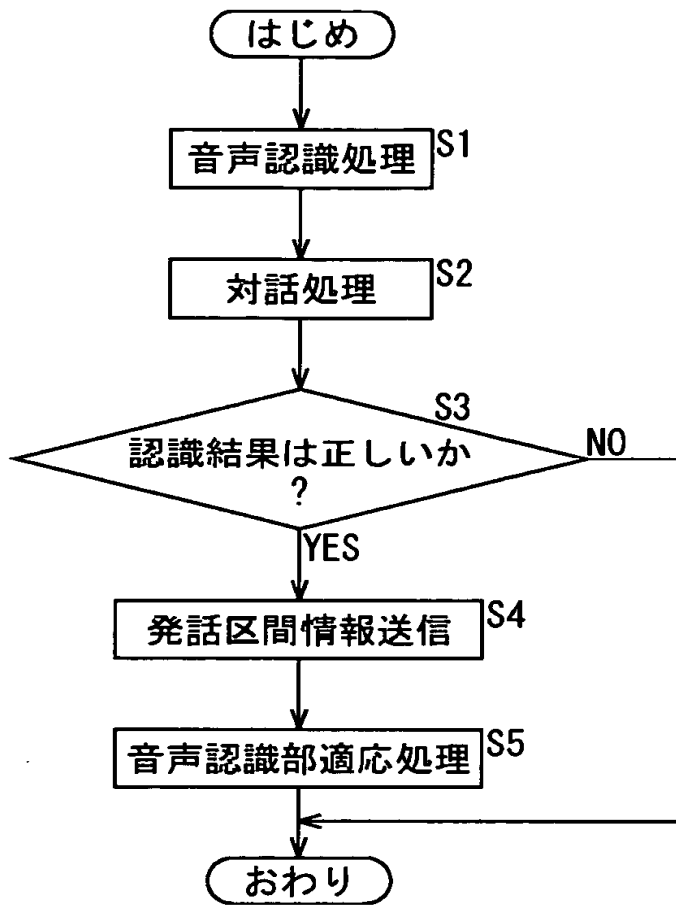
音声合成部 3

【図 6】

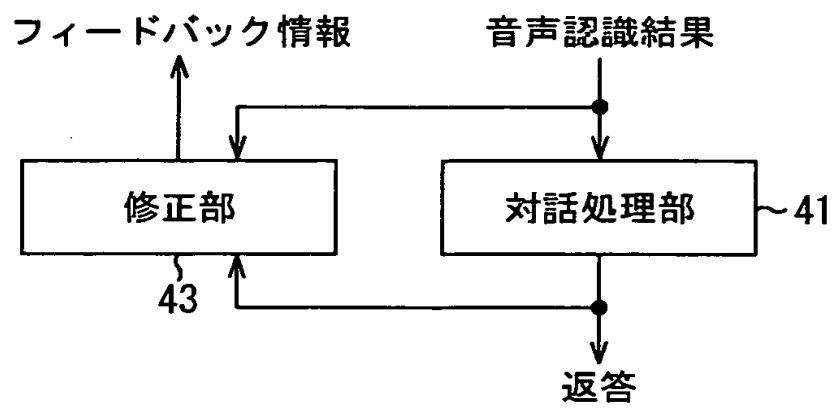


対話管理部 5

【図 7】

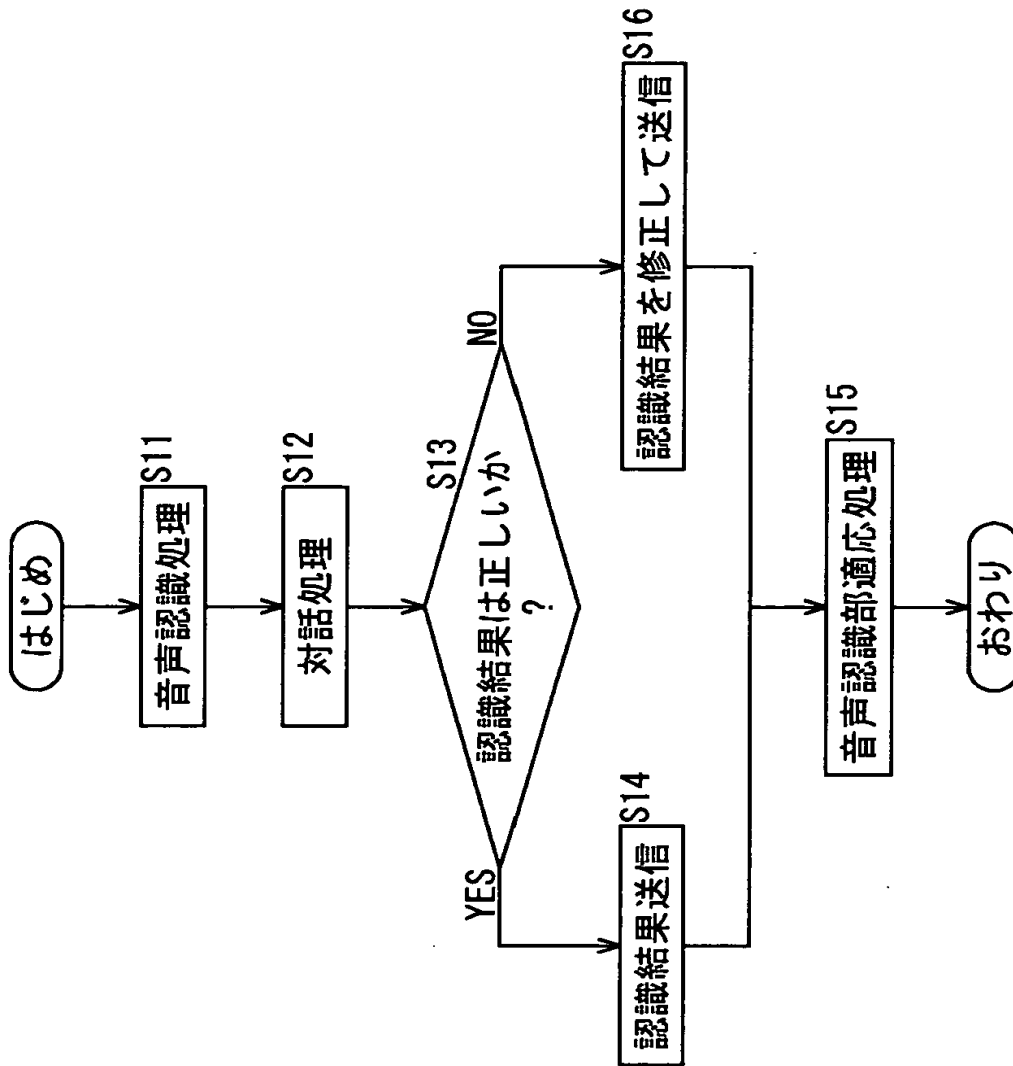


【図 8】

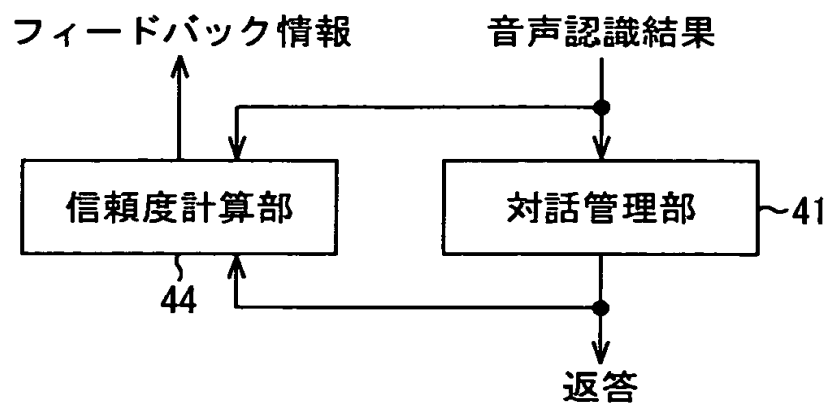


対話管理部 5

【図 9】

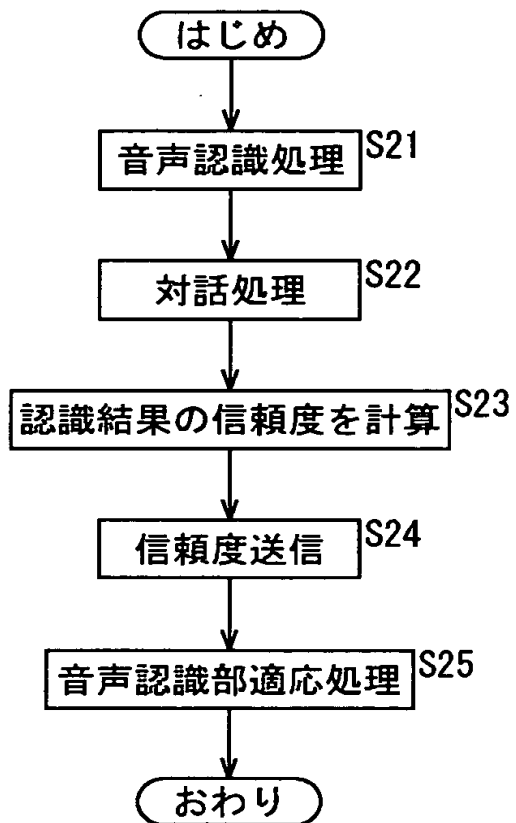


【図 1 0】

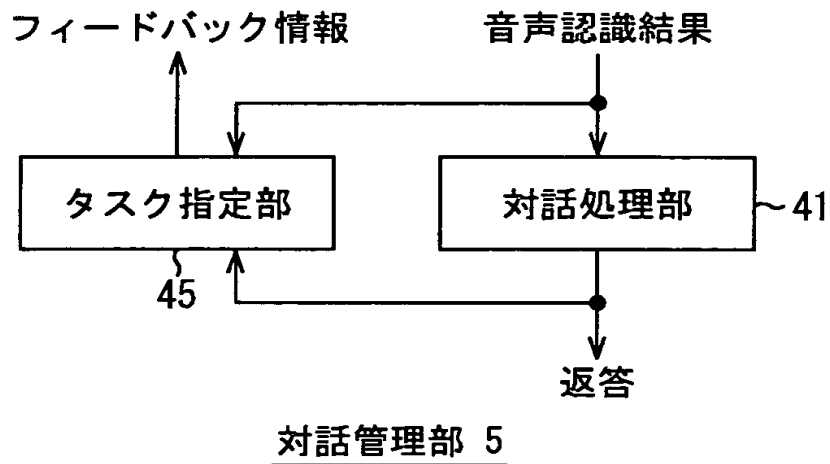


対話管理部 5

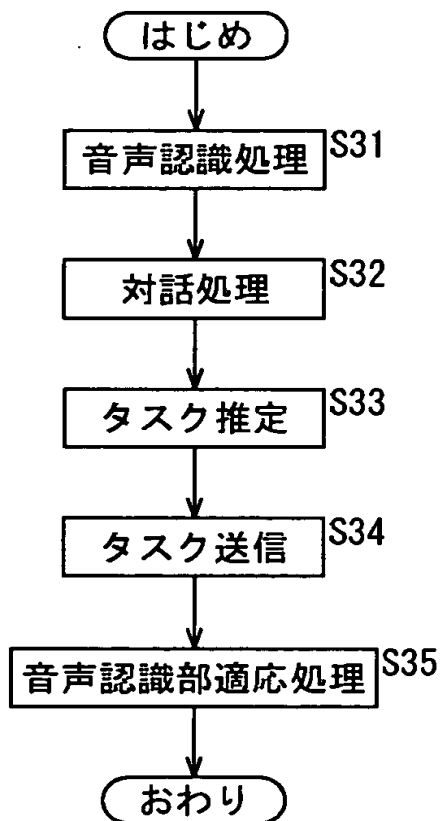
【図 1 1】



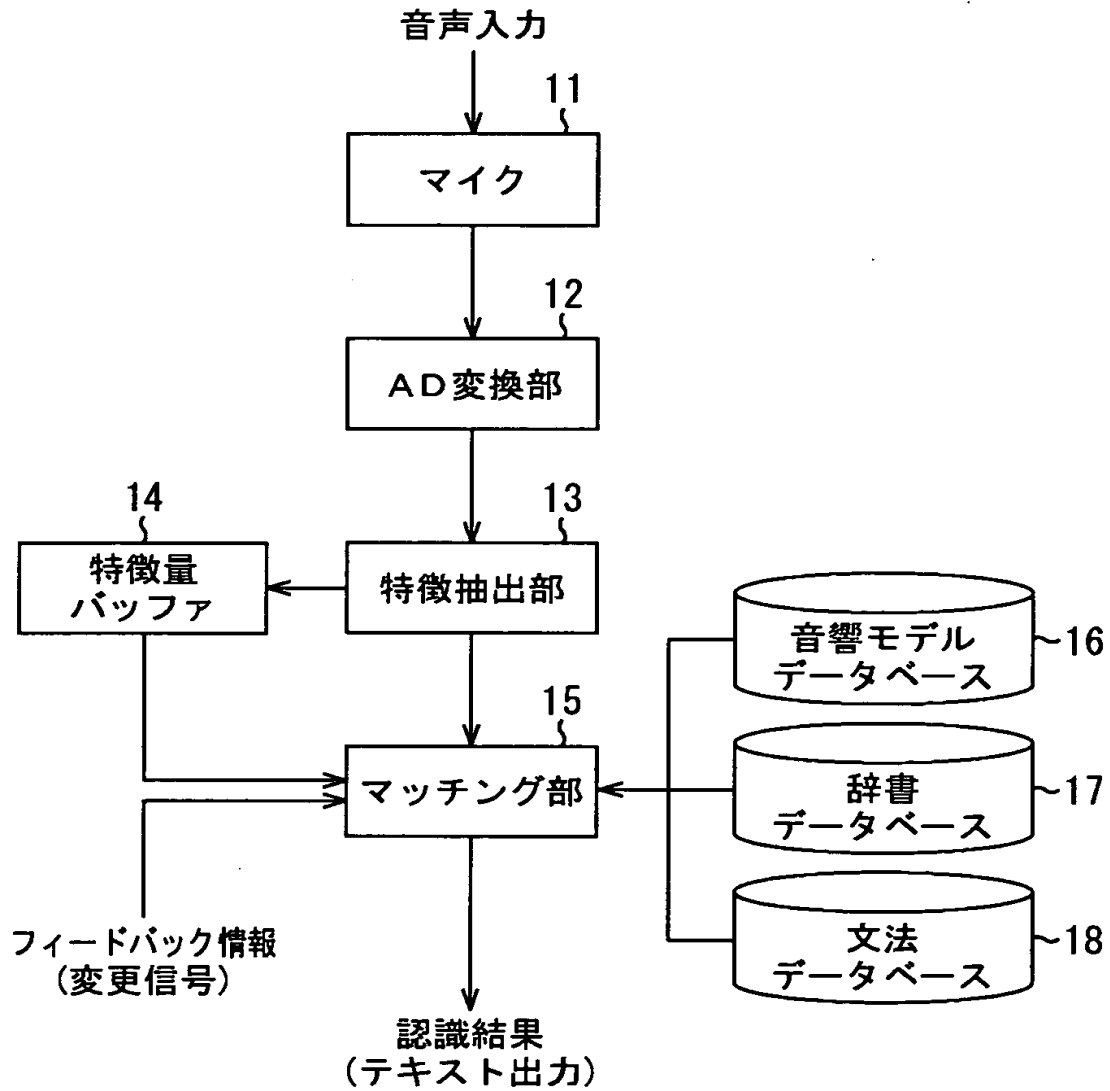
【図 1 2】



【図 1 3】

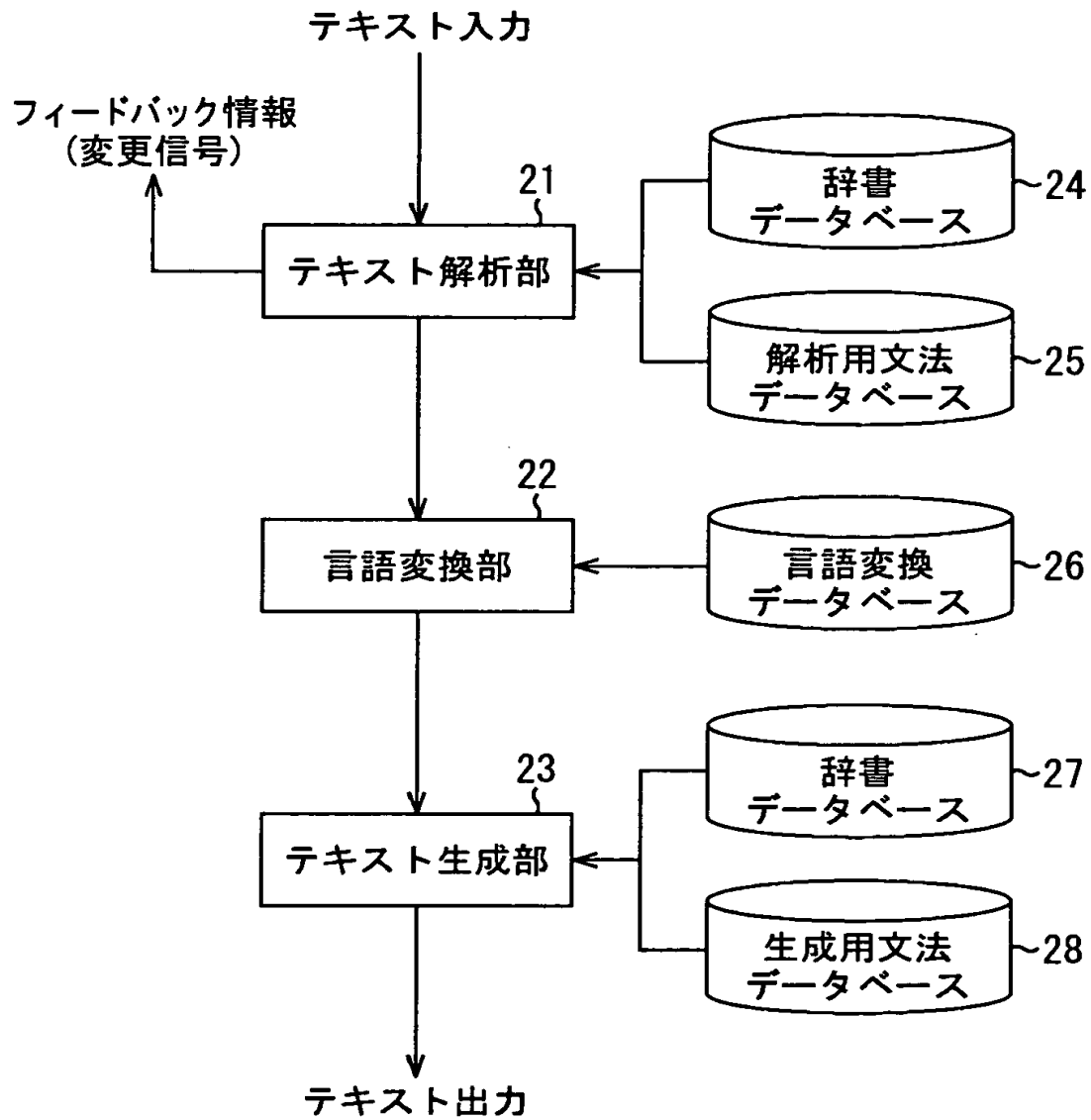


【図 1 4】



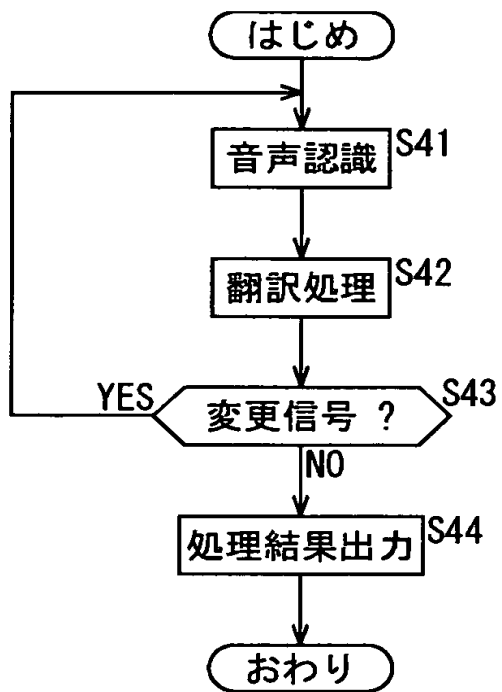
音声認識部 1

【図 1 5】

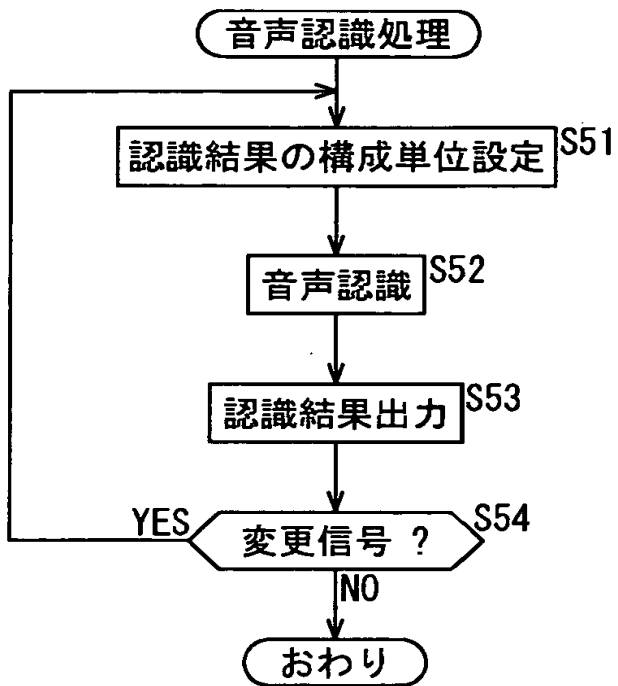


機械翻訳部 2

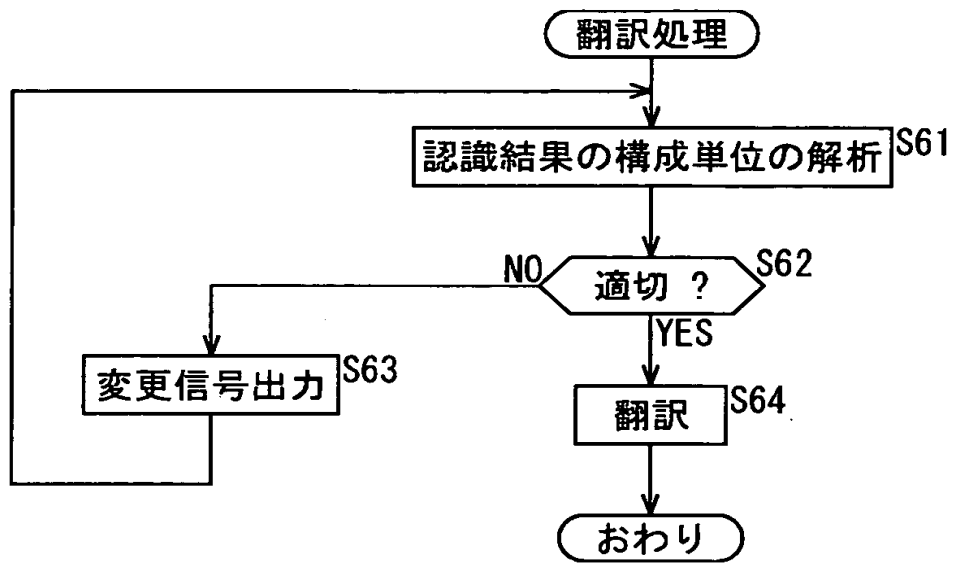
【図 1 6】



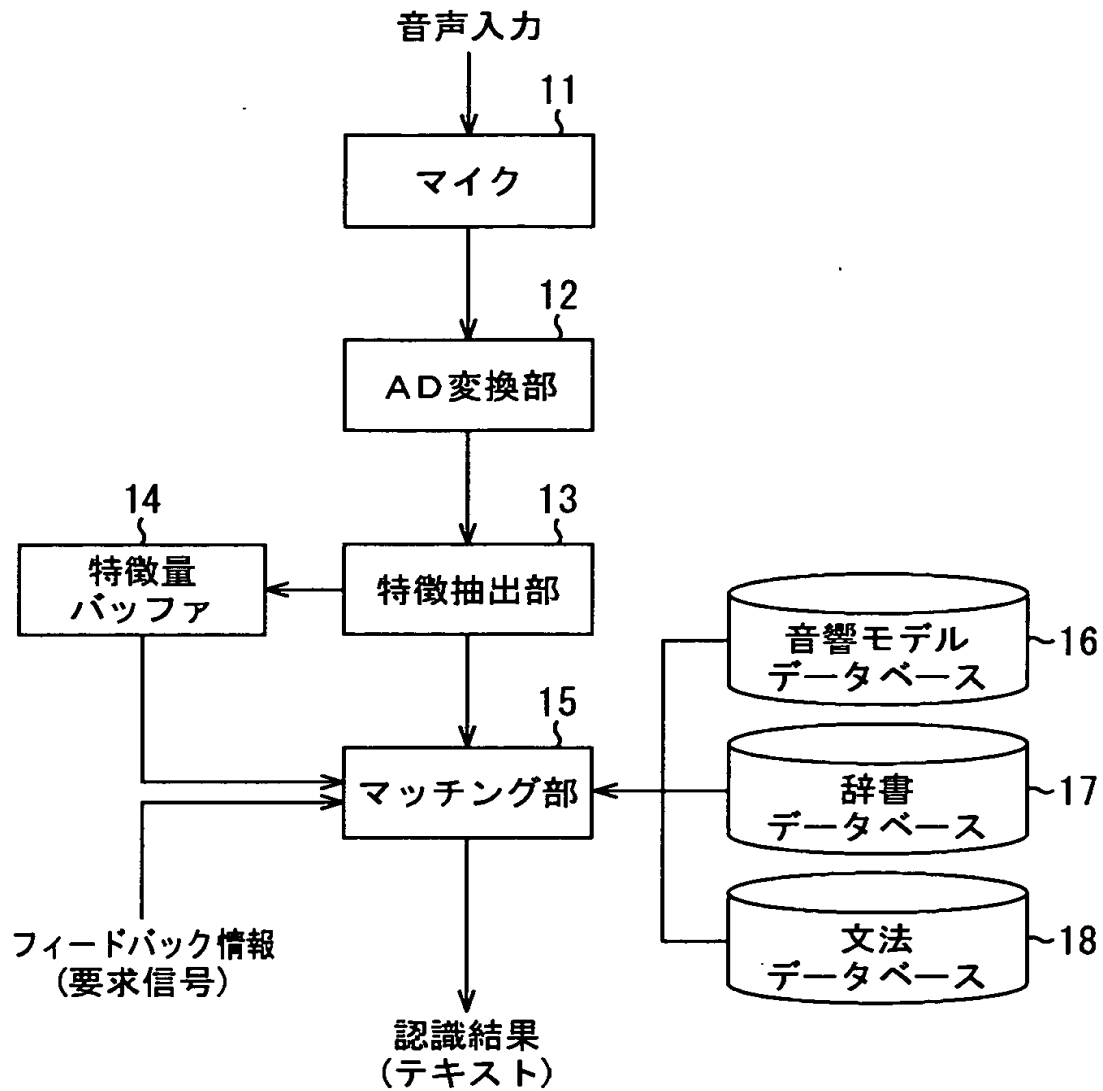
【図 1 7】



【図 1 8】

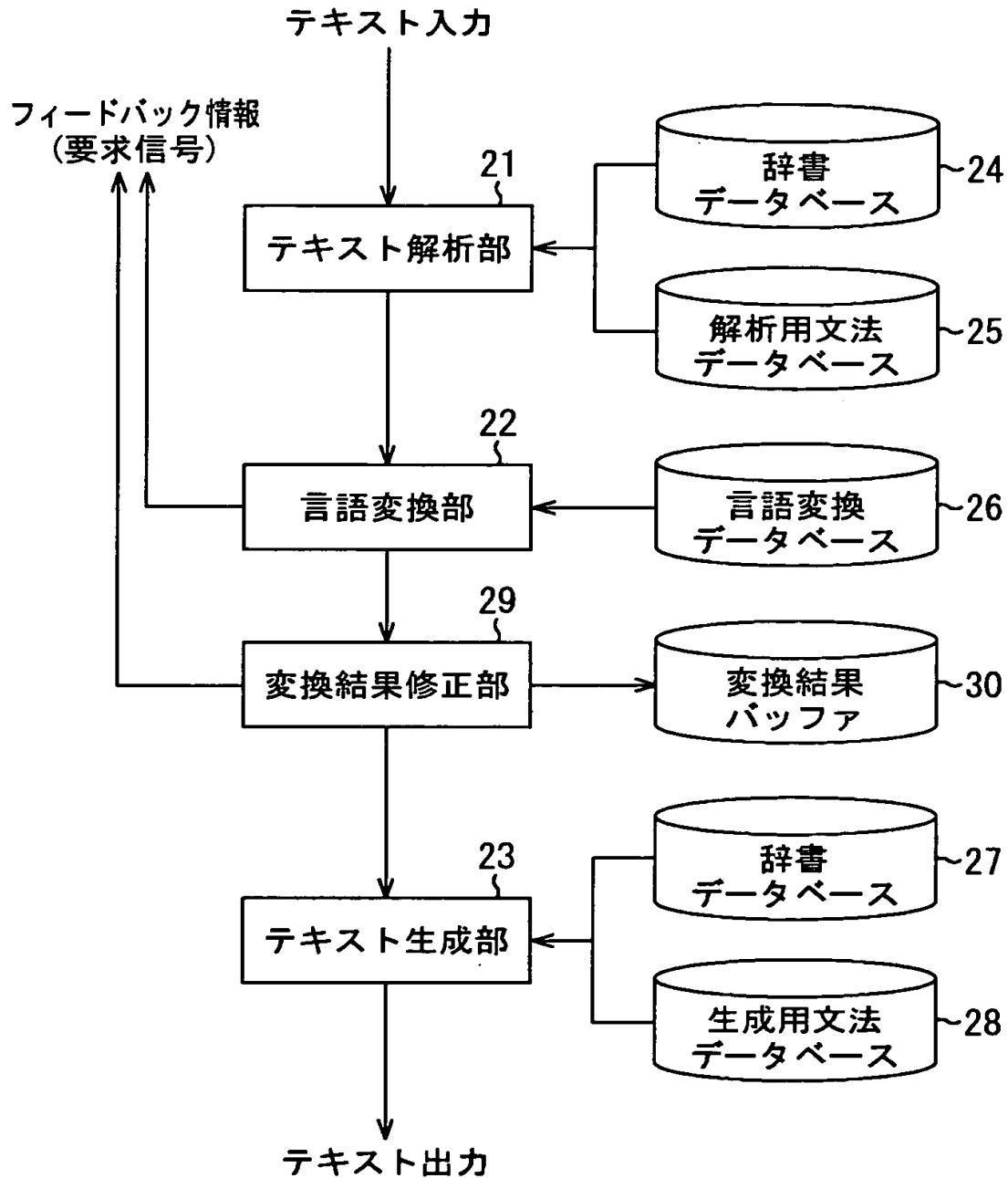


【図 1 9】



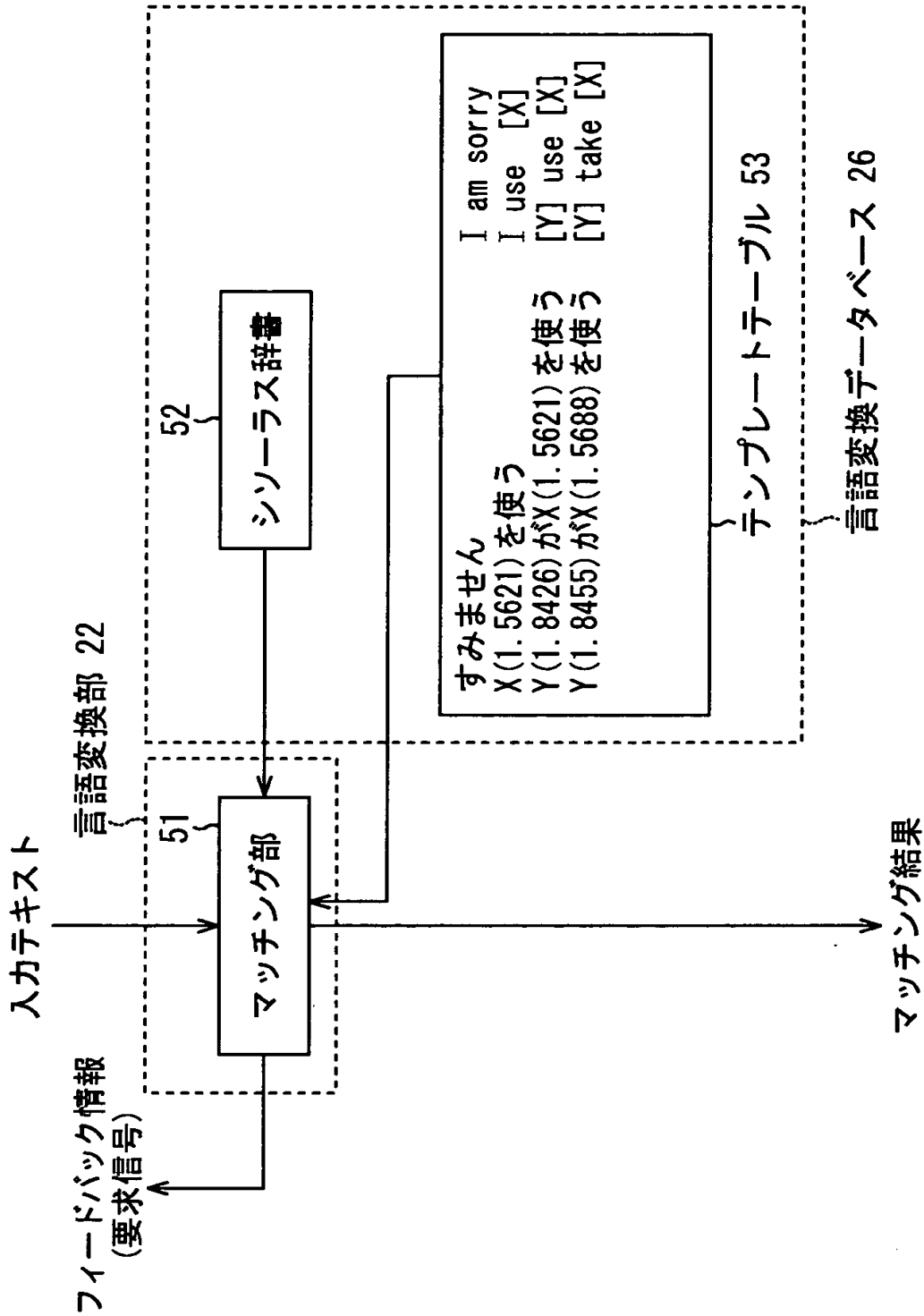
音声認識部 1

【図 2 0】

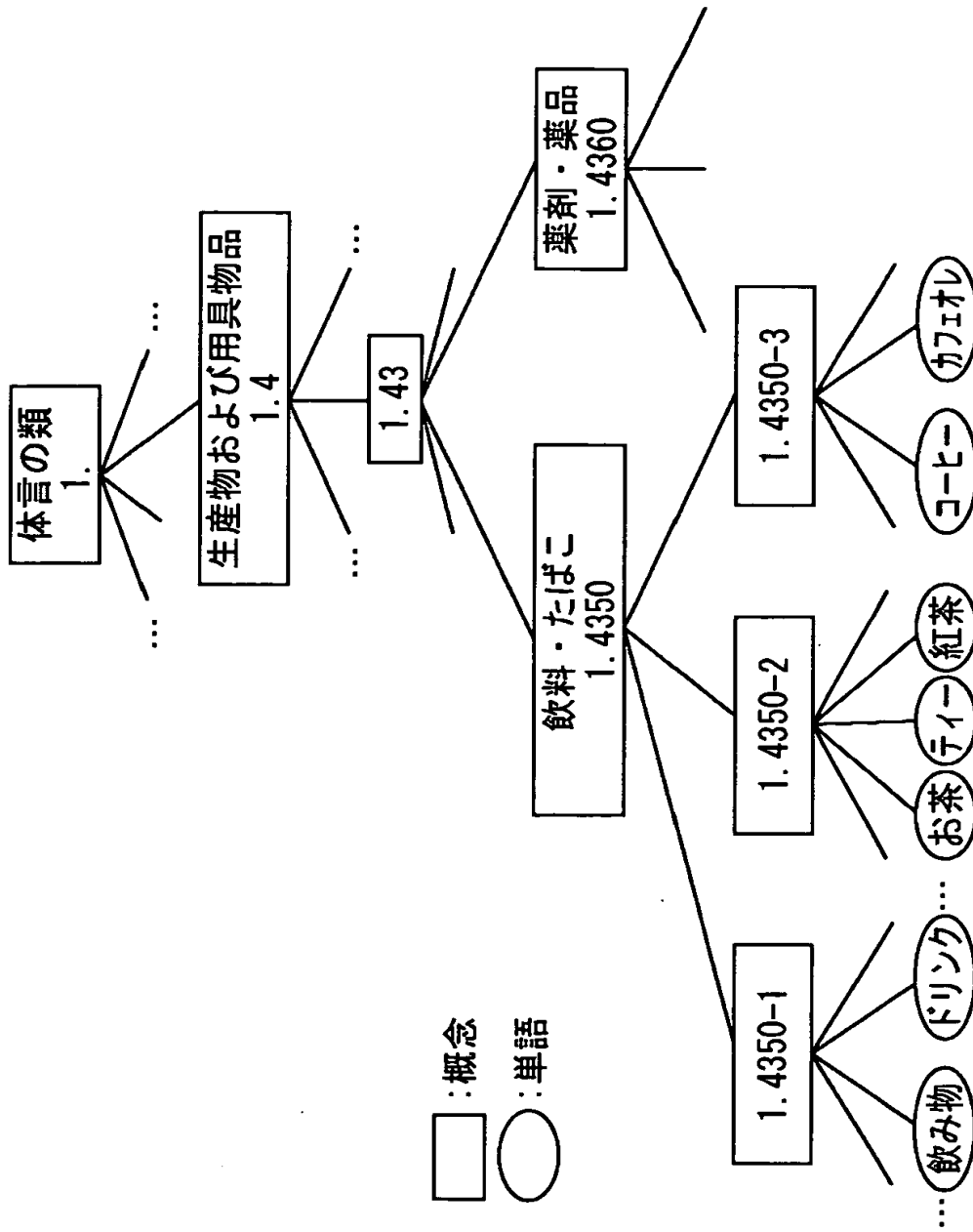


機械翻訳部 2

【図 2 1】

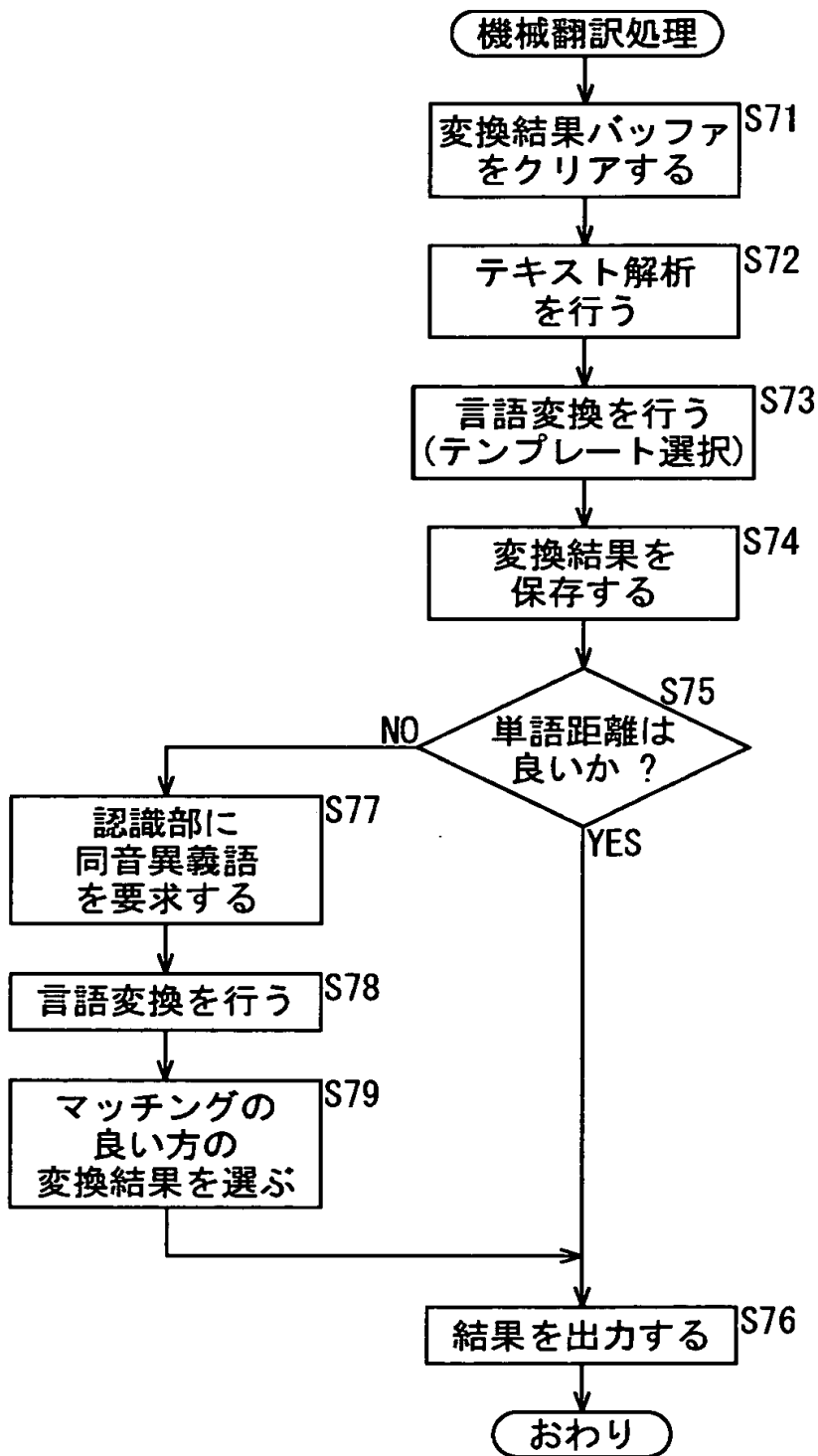


【図 2 2】

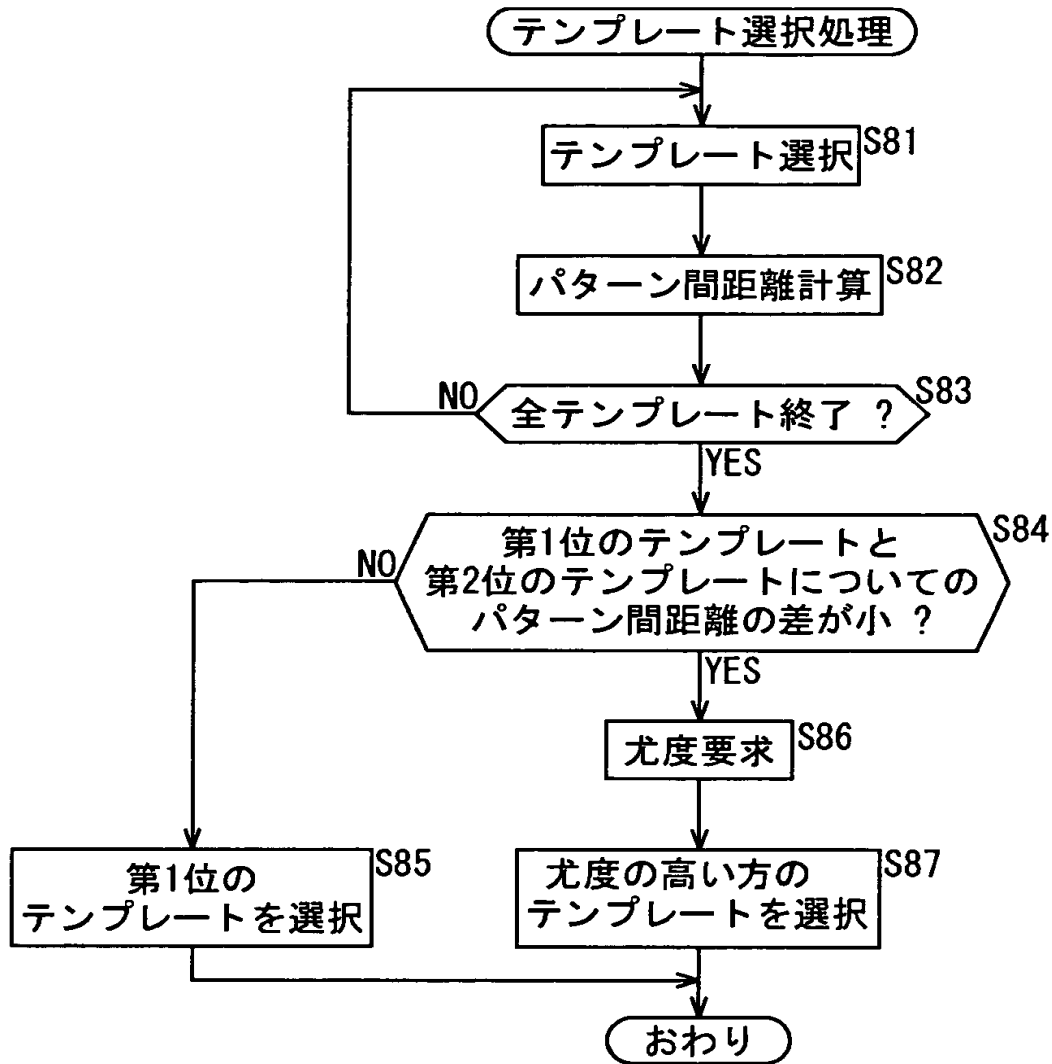


シソーラスの例(分類語い表より)

【図 2 3】



【図 2 4】



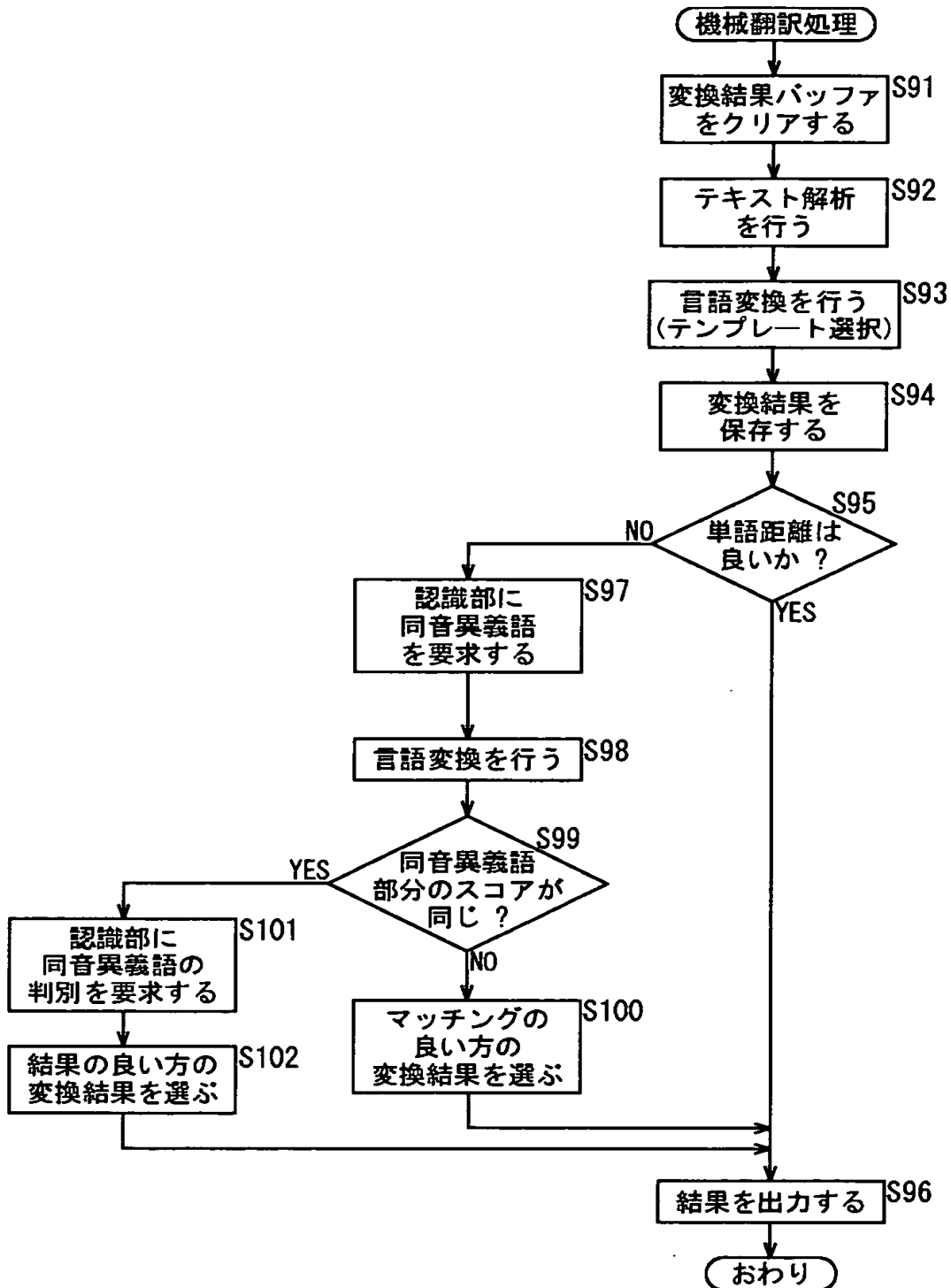
【図 2 5】

橋が：はし

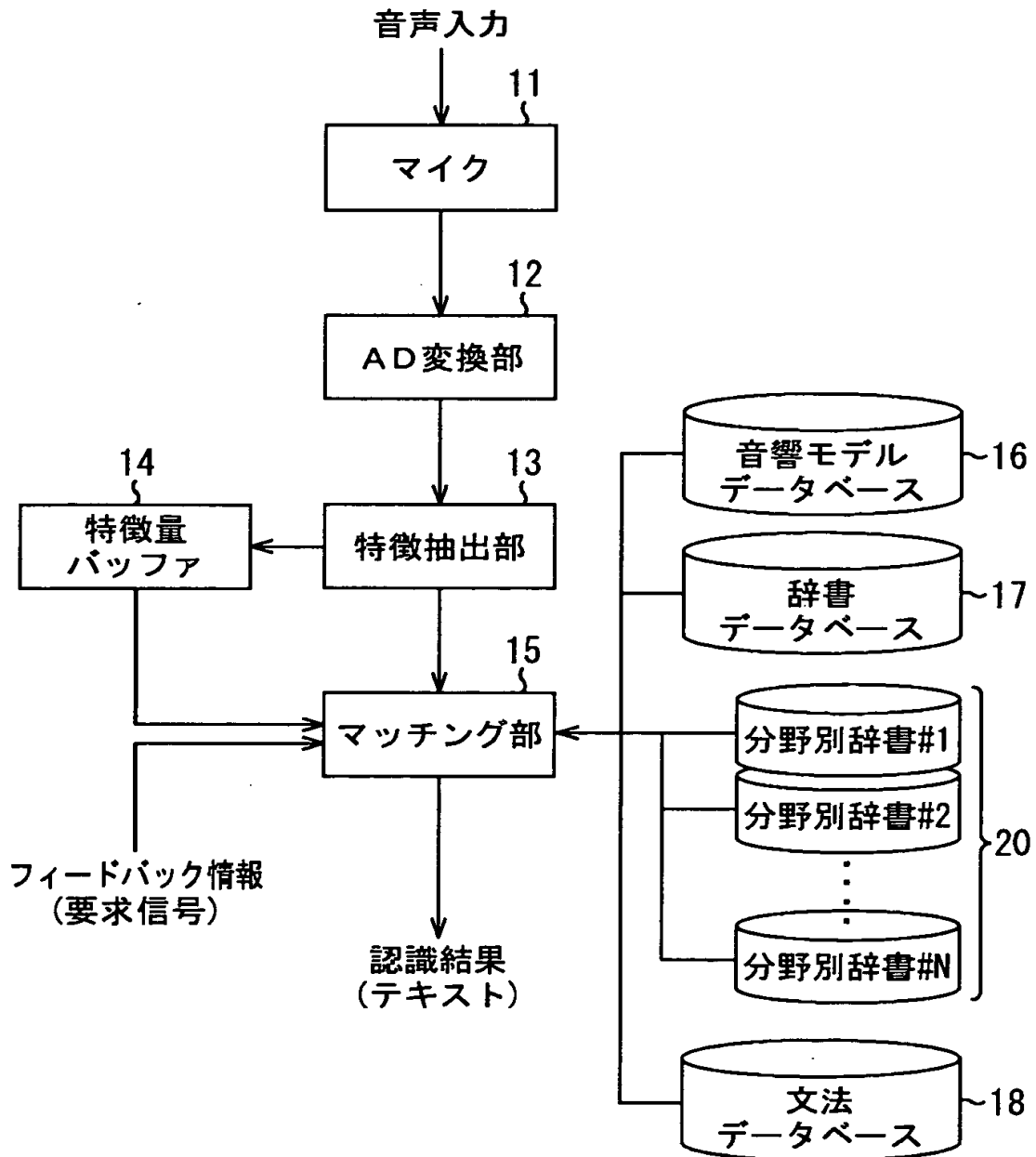
端が：はし

箸が：はし

【図 2 6】

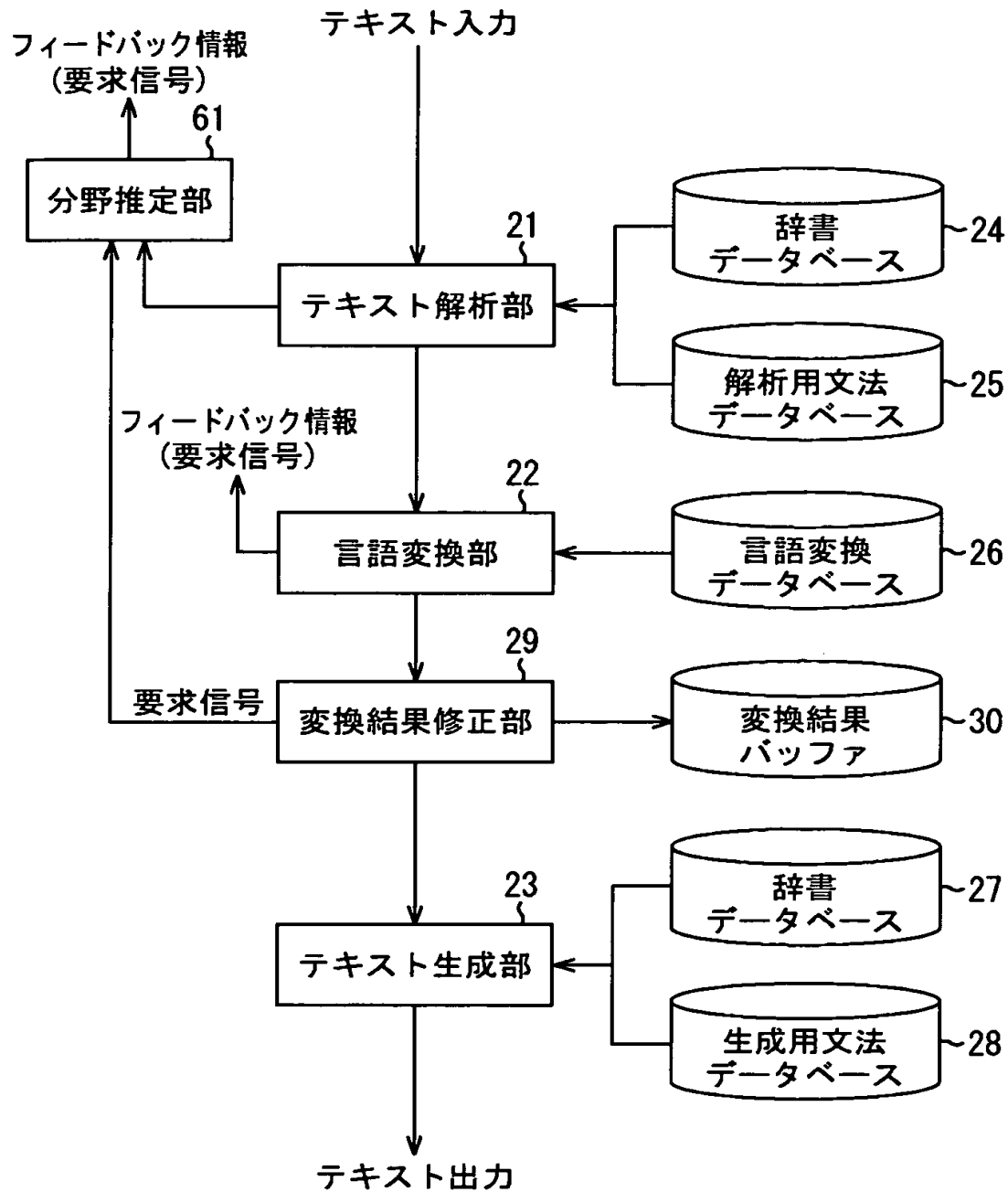


【図 2 7】



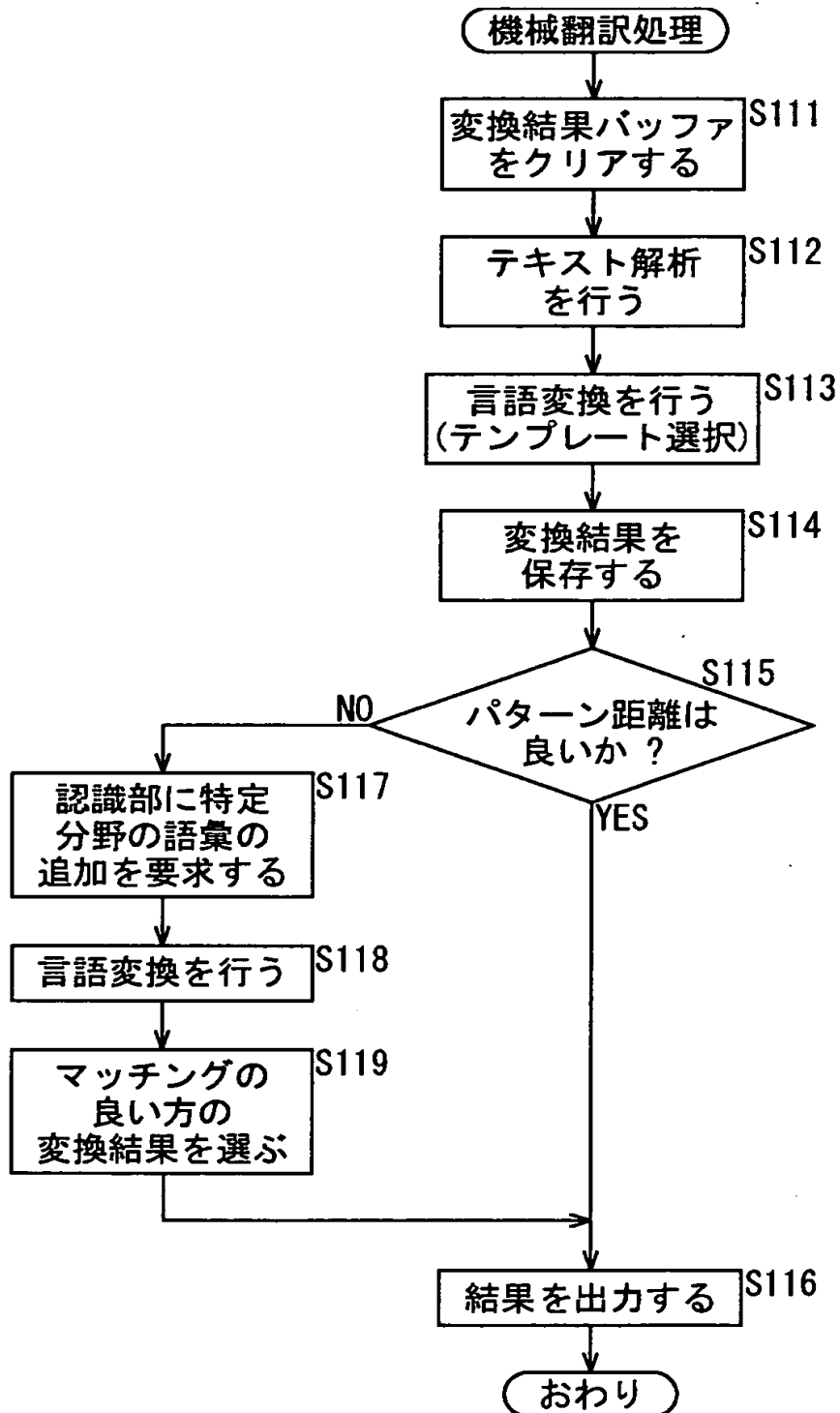
音声認識部 1

【図 2 8】



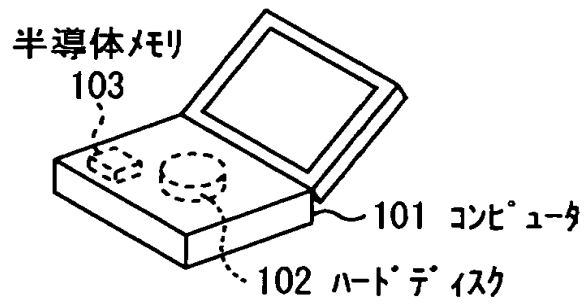
機械翻訳部 2

【図 2 9】

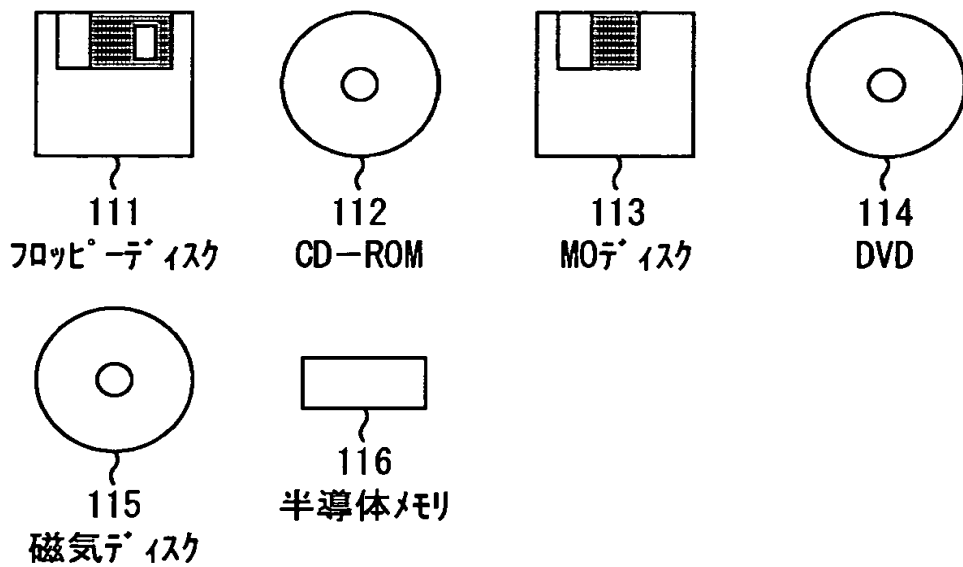


【図 3 0】

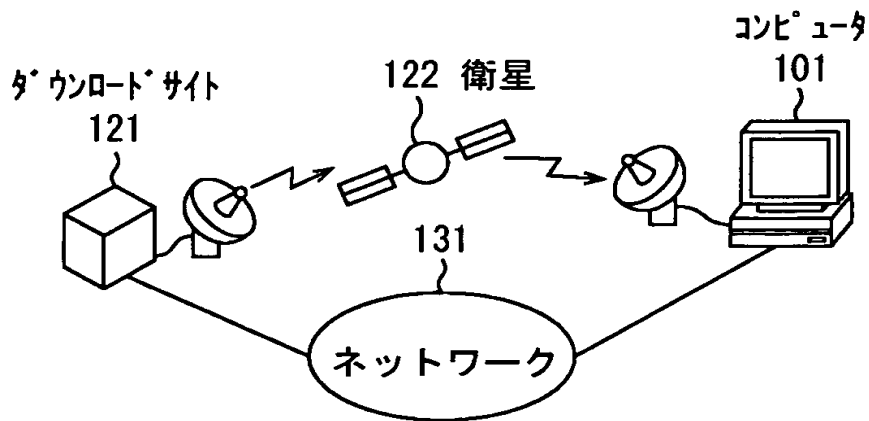
(A)



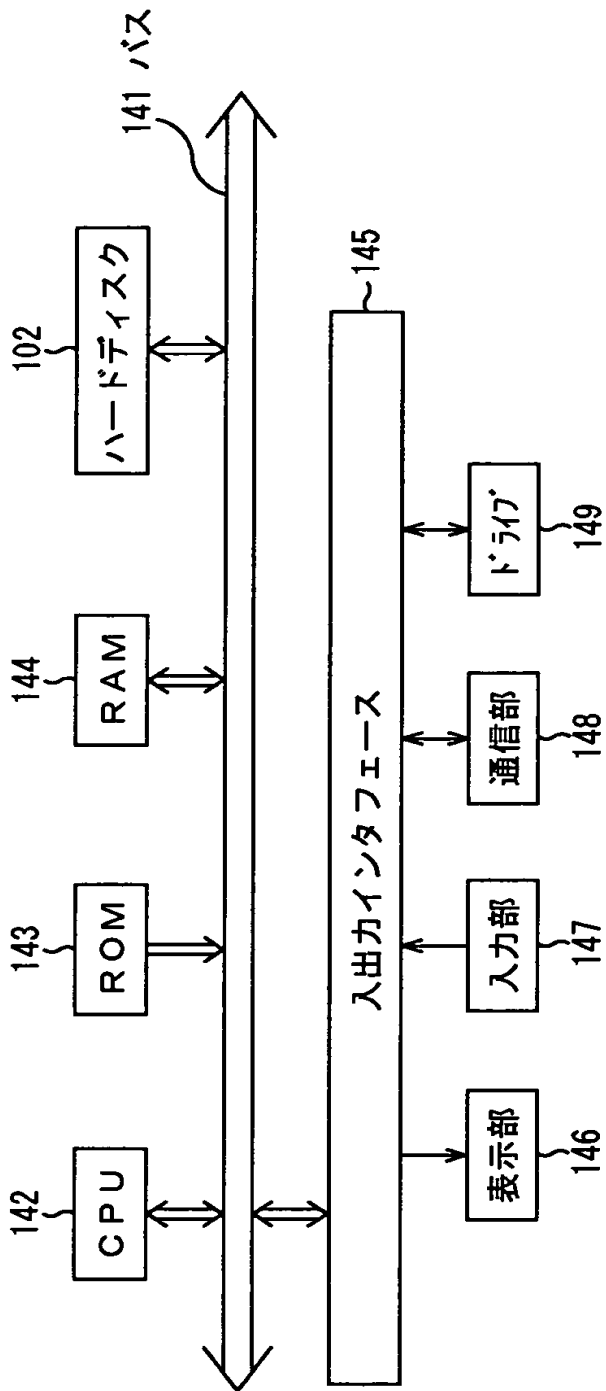
(B)



(C)



【図 3 1】



コンピュータ 101

【書類名】 要約書

【要約】

【課題】 音声認識に用いるモデルの適応を、ユーザに負担をかけることなく、かつ精度良く行う。

【解決手段】 音声認識部 1 では、そこに入力される音声、例えば、HMM (Hidden Markov Model) 等のモデルに基づいて音声認識され、その音声認識結果が、対話管理部 5 に供給される。対話管理部 5 では、音声認識部 1 からの音声認識結果に対する返答が生成される。この場合において、対話管理部 5 は、音声認識部 1 からの音声認識結果と、その返答とに基づいて、音声認識結果が正しい音声区間を検出し、その音声区間を、音声認識部 1 にフィードバックする。音声認識部 1 では、対話管理部 5 からフィードバックされる音声区間の音声と、その音声認識結果に基づいて、モデルのオンライン適応が行われる。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 2 1 8 5]

1. 変更年月日 1 9 9 0 年 8 月 3 0 日

[変更理由] 新規登録

住 所 東京都品川区北品川 6 丁目 7 番 3 5 号

氏 名 ソニー株式会社